PROTEIN FOLDING ENZYMES:
PROTEIN DISULFIDE ISOMERASE AND PROLYL 4-HYDROXYLASE

by

Elizabeth Alexander Kersteen

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

(Biochemistry)

at the

UNIVERSITY OF WISCONSIN–MADISON

2005

Committee's Page. This page is not to be hand-written except for the signatures

A dissertation entitled

Protein Folding Enzymes:
Protein Disulfide Isomerase and 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

Elizabeth Alexander Kersteen

Date of Final Oral Examination: March 24, 2005

Month & Year Degree to be awarded: **December** **May** 2005 **August**

* * *

Approval Signatures of Dissertation Committee

Richard R Burgess

Alan Attie

Silvia Cavagnero

Signature, Dean of Graduate School

Martin Cadwallader/EH

Committee's Page. This page is not to be hand-written except for the signatures

PROTEIN FOLDING ENZYMES: PROTEIN DISULFIDE ISOMERASE AND PROLYL 4-HYDROXYLASE

Elizabeth Alexander Kersteen

Under the supervision of Professor Ronald T. Raines

At the University of Wisconsin–Madison

The efficient formation of native protein structure requires an ensemble of chaperones and enzymes that improve the rate of protein conformational changes. Specific chemical modifications are catalyzed by protein folding enzymes. The native disulfide bond conformation is attained by disulfide bond formation, reduction and isomerization, processes catalyzed by protein disulfide isomerase (PDI) in the endoplasmic reticulum. Collagen, the most abundant animal protein, requires hydroxylation of proline residues for physiological stability. Prolyl 4-hydroxylase (P4H) post-translationally converts specific proline residues in collagen to 4(*R*)-hydroxy-L-proline (Hyp). This dissertation describes advances in our understanding of the mechanisms of disulfide bond isomerization and prolyl hydroxylation.

The mechanism of disulfide bond isomerization by PDI is poorly understood despite decades of research. The complexity of common substrates used to study isomerization makes it difficult to decipher details of the kinetic and chemical mechanism. Chapter Two describes the development of a novel assay for disulfide bond isomerization that uses a synthetic peptide substrate with two nonnative disulfide bonds. Isomerization is monitored directly by fluorescence resonance energy transfer. Using this new assay we have analyzed the reaction mechanism of PDI and an active-site variant.

P4H is an $\alpha_2\beta_2$ tetramer in which the α subunit contains the catalytic active site and substrate binding site and the β subunit is PDI. Chapter Three describes the first system for the production of P4H tetramer in *Escherichia coli,* which will facilitate future structural and mechanistic studies of the enzyme. Further, we describe the design of an automated assay for catalysis of prolyl hydroxylation, which will be useful for mechanistic studies and screening of inhibitors.

Collagen is a triple helix composed of three strands of a repeating (Xaa-Yaa-Gly) sequence, where Xaa is often Pro and Yaa is often Hyp. Chapter Four describes the replacement of Pro and Hyp in synthetic collagen mimics with *N*-methyl-L-alanine, a nonnatural, acyclic tertiary amide. This substitution decreases the thermal stability of triple helices, suggesting that the conformational stability imparted by Pro and Hyp is not a result of their forming tertiary amides.

Acknowledgements

I am grateful for the invaluable assistance I received from many people during the preparation of this thesis. First, I would like to thank my advisor, Ron Raines, who has provided an exciting environment for scientific research. The results reported in this thesis could not have been achieved without his creativity and encouragement. His enthusiasm has helped me to overcome many hurdles along the way.

Members of team PDI in the Raines laboratory have been a constant source of creativity and support. I am particularly indebted to the encouragement of Ken Woycechowsky. Ken has been a fabulous mentor and friend to me. Our many conversations about disulfide bonds and about life will be some of my fondest memories of graduate school. Seth Barrows was the inspiriation for many of the results reported in Chapter Two. Kelly Gorres has been an exciting addition to the group. Our discussions for future research directions in the lab have been fun and I look forward to hearing about her accomplishments in the upcoming years.

The many members of the Raines lab, past and present, have all provided support and advise that has been important to my research. I would like to specifically thank Josh Higgin for his help with the development of the assay for hydroxylase activity described in Chapter Three and helpful discussions about the mechanism of prolyl 4-hydroxylase. Marcia Haigis was an enormous help in the early days in learning molecular biology techniques. She is also a great friend and a fun person to have in the lab. I have learned a great deal from Brian Miller regarding enzyme kinetics and mechanisms. I thank him for his encouragement and friendship.

The help I received from Gary Case at the Peptide Synthesis Facility at the University of Wisconsin Biotechnology Center is much appreciated. He has been an enthusiastic source of

advice and assistance. I wish to thank Amy Harms and Grzegorz Sabat at the Mass Spectrometry Facility for their generous technical advise and services.

I am grateful for the financial assistance provided by the NIH Biotechnology Training Program, the Wisconsin Alumni Research Foundation, and the Department of Biochemistry during my graduate training.

The friendship and encouragement of Jake Mulligan have been invaluable to me. My time in graduate school would not have been the same without him. Jake is the most generous and genuine person I know, and I am grateful for his support.

Finally, I would like to thank the unending encouragement of my family. My time in graduate school has seen many changes, but through it all our love and support of each other has been constant. I am so happy that my sister, Amy, and I have grown closer over the past few years even though we are geographically far apart. My mom's continual interest in what I've been doing is much appreciated, especially since it is so difficult to explain. My dad is at least partially responsible for nurturing my interest in science through all of those intricate grade school science projects we did together. I thank both of my parents for encouraging my individuality and creativity.

Table of Contents

Abstract i

Acknowledgements iii

Table of Contents v

List of Tables vii

List of Figures viii

List of Abbreviations x

Chapter One

Introduction 1

- 1.1 Protein Folding in the Cell 2
- 1.2 Dithiol Oxidation in the Cell 8
- 1.3 Protein Disulfide Isomerase 9
- 1.4 Small Molecule Mimics of PDI 19
- 1.5 Prolyl 4-Hydroxylase 23
- 1.6 Collagen 31
- 1.7 A Need for More Mechanistic Understanding 37

Chapter Two

Catalysis of Protein Disulfide Bond Isomerization 58

- 2.1 Abstract 59
- 2.2 Introduction 60
- 2.3 Materials and Methods 64
- 2.4 Results 75
- 2.5 Discussion 80

Chapter Three

Production of Human Prolyl 4-Hydroxylase in *Escherichia coli* 92

3.1 Abstract 93

3.2 Introduction 94

3.3 Materials and Methods 97

3.4 Results 107

3.5 Discussion 113

Chapter Four

Contribution of Tertiary Amides to the Conformational Stability of Collagen Triple Helices .. 127

4.1 Abstract 128

4.2 Introduction 129

4.3 Materials and Methods 130

4.4 Results and Discussion 132

4.5 Conclusion 135

Appendix I

Phas-I as a Substrate for Human Prolyl 4-Hydroxylase 138

AI.1 Introduction 139

AI.2 Materials and Methods 140

AI.3 Results 141

References 145

List of Tables

Table 1.1 Properties of oxidoreductases 46

Table 1.2 Properties of active-site variants of thioredoxin 47

Table 1.3 Properties of active-site variants of protein disulfide isomerase 48

Table 1.4 Properties of small-molecule dithiol catalysts of disulfide isomerization 49

Table 1.5 The effect of peptide chain length on K_m for binding to type I and type II prolyl 4-hydroxylase 50

Table 2.1 Catalysis of disulfide bond isomerization by PDI 85

Table 3.1 Purification of full-length tetrameric prolyl 4-hydroxylase 118

Table 3.2 Kinetic parameters for prolyl hydroxylation by P4Hα(I)235–534/β and full-length P4H 119

Table 3.3 Comparison of HPLC-based assay to $[^{14}C]CO_2$-release assay 120

Table 4.1 Values of T_m for synthetic $(Pro\text{-}Hyp\text{-}Gly)_3(Xaa\text{-}Yaa\text{-}Gly)(Pro\text{-}Hyp\text{-}Gly)_3$ triple helices 136

Table AI.1 Results from MALDI–MS of trypsin-digested rat Phas-I 142

List of Figures

Figure 1.1 The pathway of native disulfide formation in the lumen of the ER 51

Figure 1.2 PDI is composed of four thioredoxin-like domains 52

Figure 1.3 Mechanism of disulfide isomerization 53

Figure 1.4 Reaction catalyzed by prolyl 4-hydroxylase 54

Figure 1.5 Putative mechanism for prolyl hydroxylation by P4H 55

Figure 1.6 Proposed mode of peptide substrate binding to P4H 56

Figure 1.7 Structure of the triple helix of a collagen peptide mimic 57

Figure 2.1 The mechanism of PDI-catalyzed isomerization 86

Figure 2.2 Tachyplesin I 87

Figure 2.3 Scheme for the synthesis of dns-sTI 88

Figure 2.4 Fluoresence properties of dns-nTI and dns-sTI 89

Figure 2.5 Fluorescence emission intensity of dns-nTI as a function of reduction potential. 90

Figure 2.6 Isomerization of dns-sTI by PDI 91

Figure 3.1 Reaction catalyzed by prolyl 4-hydroxylase 121

Figure 3.2 Map of plasmid pBK1.PDI1.P4H7 122

Figure 3.3 10% SDS–PAGE analysis of the production and purification of P4H from pBK1.PDI1.P4H5 123

Figure 3.4 Native PAGE analysis of purified truncated P4H 124

Figure 3.5 10% SDS–PAGE analysis of the production and purification of recombinant human P4H from pBK1.PDI1.P4H7 125

Figure 3.6 HPLC-based assay for P4H activity 126

Figure 4.1 Relative conformational stability of (Pro-Hyp-Gly)$_3$(Xaa-Yaa-Gly)(Pro-Hyp-Gly)$_3$ triple helices 137

Figure AI.1 Sequence of rat Phas-I 143

Figure AI.2 10% SDS–PAGE analysis of recombinant rat Phas-I 144

List of Abbreviations

A	absorbance
Acm	acetamidomethyl
AgOTf	silver trifluoromethanesulfonate
ANS	anthocyanidin synthase
ATF6	activating transcription factor 6
ATP	adenosine 5′-triphosphate
AtsK	alkylsulfatase from *Pseudomonas putida*
BiP	immunoglobulin heavy chain binding protein
BMC	(±)-trans-1,2-bis(mercaptoacetamido)cyclohexane
BPTI	bovine pancreatic trypsin inhibitor
CD	circular dichroism
cDNA	complementary deoxyribonucleic acid
CFTR	cystic fibrosis transmembrane conductance regulator
c.p.s.	counts per second
CS	clavaminate synthase
CXC	Cys-Xaa-Cys
CXXC	Cys-Xaa-Xaa-Cys
DAOCS	deacetoxycephalosporin C synthase
DFT	density functional theory
DIPEA	*N*,*N*-diisopropylethylamine
DMF	dimethylformamide

DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
dns	dansyl, 5-(dimethylamino)naphthalene-1-sulfonyl
dns-nTI	dansyl native tachyplesin I
dns-sTI	dansyl scrambled tachyplesin I
DsbDα	*N*-terminal domain of DsbD
DTNB	5, 5′-dithiobis(2-nitrobenzoic acid)
DTT	D,L-dithiothreitol
E°	reduction potential
$E^{\circ\prime}$	standard reduction potential
EDS	Ehlers-Danlos syndrome
EDT	2-ethanedithiol
EDTA	ethylenediaminetetraacetic acid
eIF4E	eukaryotic initiation factor 4E
EPR	electroparamagnetic resonance
ER	endoplasmic reticulum
ERAD	endoplasmic reticulum-associated degradation
Ero1p	endoplasmic reticulum oxidoreductin 1 protein
Ero1-Lα	endoplasmic reticulum oxidoreductin 1-Like protein α isoform
Ero1-Lβ	endoplasmic reticulum oxidoreductin 1-Like protein β isoform
ESI	electrospray ionization
Eug1p	endoplasmic reticulum protein unnecessary for growth 1

ε	extinction coefficient
F	fluorescence emission intensity
F_0	initial fluorescence emission intensity
FAD	flavin adenine dinucleotide
FIH-1	factor-inhibiting hypoxia-inducible factor 1
Flp	4(*R*)-fluoro-L-proline or (2*S*, 4*R*)-4-fluoroproline
F_{max}	maximal fluorescence emission intensity
Fmoc	fluorenylmethoxycarbonyl
$F_{\mathrm{P(SH)_2}}$	formal concentration of a reduced molecule
$F_{\mathrm{PS_2}}$	formal concentration of an oxidized molecule
FPLC	fast performance liquid chromatography
FRAP/mTOR	FKBP12-rapamycin associated protein/mammalian target of rapamycin
FRET	fluorescence resonance energy transfer
$\mathcal{F}$	Faraday constant
GFP	green fluorescent protein
gor	glutathione reductase gene
GPI	glycophosphatidylinositol
Grx	glutaredoxin reductase
GSH	reduced glutathione
GSSG	oxidized glutathione
h	hour
h	Hill coefficient of cooperativity

HBTU	*O*-(benzotriazol-1-yl)-*N,N,N′,N′*-tetramethyluronium hexafluorophosphate
HCl	hydrochloric acid
HDEL	histidine-aspartate-glutamate-leucine
HIV-1	human immunodeficiency virus 1
HOBT	1-hydroxybenzotriazole
HPLC	high-performance liquid chromatography
Hsp60	heat shock protein 60
Hsp70	heat shock protein 70
Hsp90	heat shock protein 90
Hyp	4(*R*)-hydroxy-L-proline or (2*S*,4*R*)-4-hydroxyproline
IPTG	isopropyl-1-thio-β-D-galactoside
IRE1	inositol-requiring enzyme 1
K_a	acid dissociation constant
k_{cat}	first-order enzymatic rate constant
K_d	equilibrium dissociation constant
kDa	kilodalton
K_{eq}	equilibrium constant
K_m	Michaelis constant
K_r	slope obtained from a plot of the log of the relative motility versus gel concentration
KDEL	lysine-aspartate-glutamate-leucine
LB	Luria–Bertani
LC/MS	liquid chromatography/mass spectrometry
MALDI	matrix-assisted laser desorption/ionization

MALDI–TOF	matrix-assisted laser desorption/ionization time of flight
meAla	*N*-methyl-L-alanine
min	minute
mRNA	messenger ribonucleic acid
MS	mass spectrometry
MW	molecular weight
NaCl	sodium chloride
$NADP^+$	β-nicotinamide adenine dinucleotide phosphate, oxidized form
NADPH	β-nicotinamide adenine dinucleotide phosphate, reduced form
Nleu	*N*-isobutylglycine
NMA	*N*-methylmercaptoacetamide
NMR	nuclear magnetic resonance
nTI	native tachyplesin I
OD	optical density
P4H	collagen prolyl 4-hydroxylase
P4Hα(I)	isoform I of the α subunit of human collagen prolyl 4-hydroxylase
PAGE	poly(acrylamide) gel electrophoresis
PCR	polymerase chain reaction
PDI	protein disulfide isomerase
PEI	poly(ethylenimine)
PERK	double-stranded RNA-activated protein kinase-like ER kinase
Phas-I	phosphorylated heat- and acid-stable protein
pK_a	log of the acid dissociation constant

poly(C)	poly(cytidylic acid)
PP-II	polyproline II
proCPY	procarboxypeptidase Y
rbs	ribosome binding site
R	gas constant
R_f	relative motility
RNA	ribonucleic acid
RNase A	ribonuclease A
RP–HPLC	reversed-phase high-performance liquid chromatography
rRNase A	reduced ribonuclease A
s	second
Sar	sarcosine, *N*-methylglycine
SDS	sodium dodecyl sulfate
SH3	Src homology 3
sRNase A	scrambled ribonuclease A
sTI	scrambled tachyplesin I
TauD	taurine:α-ketoglutarate dioxygenase
TB	Terrific Broth
TF	trigger factor
TFA	trifluoroacetic acid
TfdA	2,4-dichlorophenoxyacetate/α-ketoglutarate dioxygenase
TI	tachyplesin I
TIS	triisopropylsilane

T	temperature
T_m	temperature at the midpoint of the thermal denaturation curve
TPR	tetratricopeptide repeat
Tris–HCl	tris(hydroxymethyl)aminomethane hydrochloride
Trt	trityl
Trx	thioredoxin
trxB	thioredoxin reductase gene
UPR	unfolded protein response
UV	ultraviolet
V_{max}	maximal velocity
Xaa	any amino acid
Yaa	any amino acid

Chapter One

Introduction

Portions of this chapter were published as:

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.

1.1 Protein Folding in the Cell

Folding of proteins into specific three-dimensional structures is necessary for their biological activity. Nearly 50 years ago, Anfinsen demonstrated that the information required to fold a protein into its native structure is contained entirely within its amino acid sequence (Sela *et al.*, 1957; Anfinsen, 1973). Following reduction and denaturation, the small protein ribonuclease A (RNase A) was able to fold into its active conformation *in vitro* without any external assistance (Sela *et al.*, 1957). Researchers quickly began to ask how a protein sequence reaches the native state. It was apparent that a protein would not be able to randomly sample all possible conformations on the way to a native structure. Levinthal was the first to propose that proteins must fold by pathways that proceed through specific folding intermediates to a native structure dictated by the most thermodynamically stable state (reviewed in (Dill & Chan, 1997)). More recent proposals suggest that the protein folding pathway is much less defined than the more classical view suggested. Instead, proteins likely take many routes on their way down a three-dimensional energy landscape with many peaks and valleys (Baldwin, 1995). Through the use of parallel processes and ensembles of intermediates with increasingly native conformations, the most efficient rate of protein folding can be achieved.

In vitro experiments have suggested that folding begins with the formation of a folding nucleus around which the remaining polypeptide can assemble (Fersht, 2000). This initial step involves the interaction of a small number of hydrophobic and polar amino acids. Hydrophobic collapse is generally followed by the formation of secondary structural elements. Finally covalent interactions, including disulfide bonds, help to stabilize protein conformations. Although this process is fairly efficient with simple proteins, the rate of protein folding is significantly impaired as the complexity of the native protein fold increases (Plaxco *et al.*, 1998).

The protein folding situation *in vivo* has even greater complexity than that *in vitro*. Cellular conditions cannot be altered to improve folding efficiency and are often unfavorable for the process. High temperatures, molecular crowding, and a large number of nascent and misfolded proteins in the cellular medium make protein folding difficult. To combat this compromising situation, the cell has devised protein quality control mechanisms, which assist protein folding, prevent protein misfolding and aggregation, and eliminate terminally misfolded proteins. The cell accomplishes these complicated tasks with the help of a large number of molecular chaperones and protein folding enzymes.

Molecular Chaperones

A class of ubiquitous proteins termed molecular chaperones assist proper folding in the cell by binding co- and post-translationally to unfolded or partially folded polypeptides. Chaperones bind to hydrophobic regions of nonnative proteins to prevent aggregation and limit the accumulation of unproductive intermediates along the folding pathway. Through the controlled binding and release of peptide substrates, chaperones allow for efficient folding to the native protein conformation.

As soon as they begin to exit the ribosome, nascent polypeptide chains encounter various chaperones. The *Escherichia coli* trigger factor (TF) protein is one example of a ribosome-bound chaperone. Its crystal structure was recently solved and gives a unique picture of how a chaperone binds both to the ribosome and to unfolded protein substrates (Ferbitz *et al.*, 2004). The protein forms a hydrophobic cradle at the exit channel of the ribosome that is large enough for an exiting peptide to begin to fold in an environment that is protected from the cytosol. TF may also migrate away from the ribosome in order to transport partially folded protein substrates to additional chaperones or subcellular locations (Bukau *et al.*, 2000). In eukaryotes, a similar but unrelated chaperone system, the nascent chain-associated complex, also binds proteins as they exit the ribosome (Wang *et al.*, 1995).

A large ubiquitous class of chaperones termed the Hsp70s (for heat shock protein 70) assist protein folding co- and post-translationally. The Hsp70s bind unfolded nascent proteins in extended, flexible conformations through hydrophobic interactions and hydrogen bonds (Zhu *et al.*, 1996; Rüdiger *et al.*, 1997). Binding and release of substrate to Hsp70 is controlled by cycles of ATP hydrolysis with the assistance of cochaperones (reviewed in (Walter & Buchner, 2002)). The Hsp70 chaperones do not catalyze conformational changes or provide a significant microenvironment for protein folding but improve the chances for attaining the native state by preventing unproductive aggregation.

The Hsp60 chaperones (or chaperonins), on the other hand, provide an isolated environment for the post-translational folding of cytosolic proteins (reviewed in (Feldman & Frydman, 2000; Walter & Buchner, 2002)). The most characterized chaperonin is the GroEL/GroES system from *E. coli*, but homologs have been identified ubiquitously (Gutsche *et al.*, 1999). Extensive structural analysis has revealed that the oligomeric structure of GroES with GroEL forms a flexible, capped barrel structure (Braig *et al.*, 1994; Hunt *et al.*, 1996) which accommodates partially folded polypeptide substrates between 20 and 60 kDa (Ewalt *et al.*, 1997). As with the Hsp70s, chaperonins regulate binding and release through cycles of ATP hydrolysis.

Additional chaperones exist in lower abundance in both prokaryotes and eukaryotes. The Hsp90s in eukaryotes, for example, use many cofactors and appear to form relatively stable complexes with substrate proteins (Walter & Buchner, 2002). The process of protein folding in this system may be more substrate specific than it is with the more abundant Hsp70s and chaperonins. The specificity of chaperones for specific substrates is still largely unknown and with further research we may discover additional complexities to the recognition and folding of substrates by these proteins and protein complexes.

Quality Control in the Endoplasmic Reticulum

The endoplasmic reticulum (ER) of eukaryotic cells is a specialized cellular compartment for the folding of secretory proteins (reviewed in (Ellgaard & Helenius, 2003)). In addition to attaining their correct tertiary and quaternary structure, secreted proteins are also covalently modified by disulfide bond formation, *N*-linked glycosylation, and glycophosphatidylinositol (GPI)-anchor addition. A set of unique chaperones and enzymes are present in the ER to ensure the fidelity of these processes.

Secretory proteins are translocated to the ER co-translationally by *N*-terminal signal sequence recognition, and enter the lumenal space through the translocon complex (Wilkinson *et al.*, 1997; Brodsky, 1998). The abundant ER-resident Hsp70 chaperone, BiP, binds nascent polypeptides as they enter the ER (Lyman & Schekman, 1997). Like other Hsp70 family members, BiP binds extended hydrophobic regions of unfolded proteins and regulates the release of its substrates through ATP hydrolysis (Flynn *et al.*, 1989; Flynn *et al.*, 1991).

Disulfide bond formation is often the rate-determining step in the folding of a protein to its native three-dimensional structure. The ER contains a family of related oxidoreductase enzymes that catalyze the formation of native disulfide bonds. The most abundant of these, protein disulfide isomerase (PDI), is able to catalyze dithiol oxidation, disulfide bond isomerization, and disulfide bond reduction through cycles of reduction and oxidation of two independent Cys–Gly–His–Cys (CGHC) active sites. PDI is discussed in further detail in Section 1.3.

Glycoprotein folding and secretion is regulated by a system that involves a PDI-related protein, ERp57. ERp57 specifically binds calnexin and calreticulin (Oliver *et al.*, 1999), two homologous lectins in the ER, which recognize incompletely folded glycoproteins (Ou *et al.*, 1993; Peterson *et al.*, 1995). ERp57 then catalyzes disulfide bond formation and rearrangement in the bound glycoproteins. This complex is responsible for retaining *N*-linked glycoproteins in the ER until they reach their native conformation.

Under normal cellular conditions, some proteins will be unable to achieve the native state. The longer a protein remains in a misfolded structure, the lower its chances for being secreted. These terminally misfolded proteins are removed from the ER through the ER-associated degradation (ERAD) pathway, a process which involves retrotranslocation from the ER to the cytosol for degradation by the proteasome (reviewed in (Zhang & Kaufman, 2004)). ERAD helps to prevent the accumulation of misfolded protein in the ER.

Under stressful conditions to the cell, unfolded proteins have a tendency to accumulate and can be detrimental if released to the cytosol or extracellular space. The ER has developed a sensitive mechanism to recognize this situation, termed the unfolded protein response (UPR), which signals the cell to increase production of folding chaperones and enzymes (reviewed in (Rutkowski & Kaufman, 2004)). In the ER of mammals, three transmembrane sensor proteins (IRE1, ATF6 and PERK) are responsible for this regulation. These three sensors are signalled primarily by BiP. In normal situations, BiP is found bound to the lumenal domain of each of the sensor proteins, but as the protein load in the ER increases, BiP disengages from IRE1, ATF6 and PERK to help with the folding of nascent polypeptides. The sensor proteins are then free to exert their regulatory effects on transcription and translation in order to increase the protein folding capacity and decrease the rate of protein production. This highly regulated system ensures that as misfolded proteins accumulate, the ER receives more protein folding assistance in the form of chaperones and enzymes, and misfolded proteins are degraded before they can do harm.

Protein Folding and Disease

Given the complexity of cellular protein folding, it is not surprising that disease states often develop when the process encounters problems. Defects in protein folding can cause a destabilization of the native structure, changes in the kinetics of folding, incorrect or prolonged association with chaperones or folding enzymes, or the accumulation of misfolded intermediates (Thomas *et al.*, 1995).

Many diseases are the result of a single missense mutation or an amino acid deletion or insertion in a particular gene. These simple mutations commonly cause changes in the folding of the mutant proteins by changing the thermodynamic balance between the native folding pathway and a misfolding pathway. Cystic fibrosis, for example, is the result of a deletion of a Phe residue in the cystic fibrosis transmembrane conductance regulator (CFTR). The folding pathway is altered by this mutation through the destabilization of a folding intermediate (Thomas *et al.*, 1992a; Thomas *et al.*, 1992b). The mutant CFTR is associated with Hsp70 in the cytoplasm and calnexin and calreticulin in the ER, suggesting that chaperones can recognize the misfolded state but are unable to correct the defect in the folding pathway (Yang *et al.*, 1993; Pind *et al.*, 1994).

Amyloid diseases, which include Alzheimer's, Huntington's, Parkinson's and the prion diseases, are the result of accumulated protein aggregates in either the cytosol or the extracellular space (reviewed in (Dobson, 2003; Ross & Poirier, 2004)). Proteins associated with amyloid formation include both wild-type and mutant sequences, but commonly demonstrate a high tendency to misfold. Mutants may destabilize the native state by imposing both thermodynamic and kinetic barriers to obtaining the native structure. Under certain conditions these proteins manage to escape the quality control mechanisms of the cell and cause irreversible damage to cells and tissues. Aggregates in all of these diseases form an insoluble cross-β-sheet structure as the result of regions of high β-sheet propensity interacting through intermolecular main-chain hydrogen bonds. Although the role of chaperones is to prevent the formation of aggregates, in the case of amyloid diseases, the chaperones cannot keep up with the protein load. The apparent inefficiency of the quality control system may be because the formation of aggregates is too rapid or the cellular level of chaperones is insufficient.

The protein folding diseases illustrate the need for further understanding of *in vivo* protein folding pathways. As we decipher the mechanistic details of the molecular chaperones and protein folding enzymes, new methods for treating conditions that result from defects in protein folding may be revealed.

1.2 Dithiol Oxidation in the Cell

The process of native disulfide formation involves both the oxidation of dithiols to disulfides and the rearrangement of nonnative to native disulfide bonds. The recent discovery of Ero1p (endoplasmic reticulum oxidoreductin 1p) in *Saccharomyces cerevisiae* has expanded our understanding of the initial oxidation steps in the pathway of disulfide bond formation in the ER. Ero1p (and its human homologs Ero1-Lα and Ero1-Lβ (Cabibbo *et al.*, 2000; Pagani *et al.*, 2000)) is a membrane-associated ER protein that was discovered in a screen for proteins that confer resistance to dithiothreitol (DTT) when overproduced or cause sensitivity to DTT when altered (Frand & Kaiser, 1998; Pollard *et al.*, 1998).

Ero1p is an essential enzyme in yeast that oxidizes proteins in the ER (Frand & Kaiser, 1998; Pollard *et al.*, 1998). Ero1p is specific in its choice of substrates, interacting with only a few proteins (Frand & Kaiser, 1999). Most importantly, Ero1p and the folding enzyme, PDI, form a mixed disulfide *in vivo* (Frand & Kaiser, 1999). Ero1 in mammalian cells specifically oxidizes the *C*-terminal active site of PDI during the PDI-catalyzed retrotranslocation of cholera toxin (Tsai & Rapoport, 2002). Mutational analysis of Ero1p has suggested that a CXXXXC (where X refers to any amino acid) motif is responsible for oxidizing PDI and other substrates, while a distinct CXXC motif reoxidizes its catalytic site (Frand & Kaiser, 2000). This disulfide relay system is supported by recent crystollagraphic data (Gross *et al.*, 2004) which locates a Cys100–Cys105 disulfide bond in a flexible loop. This disulfide is alternatively found in a buried state in proximity to the Cys352–Cys355 active site and in a solvent accessible state that could bind to PDI or other substrate proteins. These findings, combined with evidence that PDI is involved in intermolecular disulfide bonds with its substrates (Frand & Kaiser, 1999; Molinari & Helenius, 1999), suggest a pathway for the formation of disulfide bonds in newly synthesized proteins (Figure 1.1). In the lumen of the ER, FAD-bound Ero1p oxidizes PDI (Tu *et al.*, 2000), which accumulates in its reduced form in the absence of Ero1p. Oxidized PDI then transfers oxidizing equivalents from Ero1p to reduced substrate proteins. Ero1p is then reoxidized by molecular oxygen (Tu & Weissman, 2002). The oxidative power of Ero1p, and therefore the oxidative folding cycle, is regulated by levels of free FAD in the ER (Tu & Weissman, 2002).

1.3 Protein Disulfide Isomerase

PDI (EC 5.3.4.1) was first identified almost 40 years ago as an enzyme that catalyzes the activation of reduced, and thus inactive, RNase A (Goldberger *et al.*, 1963). Anfinsen and his coworkers went on to hypothesize that the primary role of PDI is to be a "general and nonspecific catalyst for disulfide interchange in proteins containing disulfide bonds" (Givol *et al.*, 1964). From their experiments with reduced RNase A (rRNase A), they suggested that the formation of native disulfide bonds in proteins begins with the uncatalyzed air oxidation of dithiols, followed by the PDI-catalyzed isomerization of nonnative disulfide bonds. More recent studies have revealed that an ensemble of enzymes and chaperones are involved in the pathway of native disulfide formation (Zapun *et al.*, 1999; Woycechowsky & Raines, 2000). Still, PDI remains at the center of this complex biological process.

PDI is a 57-kDa ER-resident protein that catalyzes the formation, reduction, and isomerization of disulfide bonds in newly synthesized proteins (Freedman *et al.*, 1994). Independent of its catalytic activity, PDI exhibits chaperone activity by inhibiting the aggregation of unfolded proteins (Cai *et al.*, 1994; Puig & Gilbert, 1994) and is a member of at least two multimeric enzyme complexes: microsomal triglyceride transfer protein (Wetterau *et al.*, 1990; Lamberg *et al.*, 1996) and prolyl 4-hydroxylase ((Koivu *et al.*, 1987b; Pihlajaniemi *et al.*, 1987), Section 1.5)

PDI Structure

PDI has five distinct structural domains (a, a′, b, b′ and c; Figure 1.2a) as deduced from its primary and tertiary structure (Darby *et al.*, 1999). The catalytic a and a′ domains are homologous, and each contains a CGHC active-site sequence (Figure 1.2b). The b and b′ domains are also homologous to the a and a′ domains (Kemmink *et al.*, 1997; Darby *et al.*, 1999). The role of the b′ domain is to bind substrate proteins (Klappa *et al.*, 1998; Cheung &

Churchich, 1999). The cationic c domain is not required for enzymatic activity (Koivunen *et al.*, 1999) and ends with a *C*-terminal ER retention signal (KDEL in rat and human; HDEL in yeast) (Munro & Pelham, 1987). The three-dimensional structure of intact PDI is not known, but structures of the individual a, a′, b, and b′ domains reveal that each has a fold characteristic of thioredoxin (Trx; Figure 1.2b) (Kemmink *et al.*, 1996; Kemmink *et al.*, 1997).

PDI is a member of the Trx family of proteins, which is characterized by the ability to catalyze thiol–disulfide interchange reactions. Proteins of this family all share a common CXXC motif in their active site(s). The *N*-terminal cysteine residue in this motif has high reactivity at physiological pH, due in part to its location on the surface of the protein (Figure 1.2b) and its depressed pK_a value (Chivers *et al.*, 1997b). The family contains proteins with redox properties that range from strongly reducing (Trx, $E^{\circ\prime} = -0.270$ V; (Krause *et al.*, 1991)) to strongly oxidizing (DsbA, $E^{\circ\prime} = -0.120$ V; (Hennecke *et al.*, 1997; Huber-Wunderlich & Glockshuber, 1998)). The proteins are therefore equipped to perform specialized roles in a variety of cellular compartments.

Mechanism of Disulfide Isomerization

The simplest mechanism for disulfide isomerization requires only one reactive thiolate in the active site of PDI. In this scenario (Figure 1.3), the more reactive *N*-terminal cysteine residue of its CXXC motif ($pK_a = 6.7$ (Hawkins & Freedman, 1991)) provides the active-site thiolate. Nucleophilic attack of the thiolate on a nonnative disulfide bond results in the formation of a mixed-disulfide intermediate between enzyme and substrate (Darby *et al.*, 1994; Chivers *et al.*, 1998). Conformational changes in the substrate then allow a substrate thiolate to initiate disulfide rearrangements that lead, ultimately, to the formation of native disulfide bonds and the release of

PDI. Based on this mechanism, a CXXS motif should be as efficient as a CXXC motif in catalysis of disulfide isomerization. Yet, experimental evidence indicates otherwise.

Eug1p, a homolog of PDI, is a nonessential lumenal protein in *S. cerevisiae* that contains the active-site sequences CLHS and CIHS (Tachibana & Stevens, 1992). The *in vitro* isomerase activity of wild-type Eug1p is low using both reduced procarboxypeptidase Y (proCPY) and scrambled proCPY (which contains non-native disulfide bonds) as substrates (Nørgaard & Winther, 2001). In contrast, a variant of Eug1p in which the CXXS motifs are replaced with CXXC motifs is an efficient catalyst of the folding of scrambled proCPY (Nørgaard & Winther, 2001). In addition, CXXC Eug1p catalyzes the folding of scrambled proCPY and reduced proCPY folding at an almost identical rate (Nørgaard & Winther, 2001). These experiments support the earlier proposal that isomerization limits catalysis of oxidative folding by enzymes containing a CXXC motif (Givol *et al.*, 1964), and indicate that the second cysteine residue enhances catalysis.

The second active-site cysteine residue is likely to be important in rescuing trapped intermediates that can accumulate during catalysis (Figure 1.3) (Walker & Gilbert, 1997). Complex protein substrates can be slow to rearrange and become trapped in nonnative configurations that prevent successful folding. The *C*-terminal CXXC cysteine residue provides a mechanism for escape from this obstacle to efficient folding. The presence of two cysteine residues permits PDI to reduce trapped nonnative disulfide bonds and then re-oxidize the substrate to form, ultimately, the native disulfide bonds. Recent results indeed indicate that the isomerization of scrambled RNase A (sRNase A, a complex substrate with 8 randomly oxidized cysteine residues) by PDI involves cycles of reduction and reoxidation (Schwaller *et al.*, 2003).

Regulating $E^{o\prime}$ and pK_a

The ability of an oxidoreductase to be an efficient isomerase is governed by the reduction potential of its active-site disulfide bond ($E^{\circ\prime}$) and the acid dissociation constant of its nucleophilic active-site thiol (K_a) (Chivers *et al.*, 1996). For efficient substrate turnover and regeneration of catalyst, an isomerase must strike a balance between the dithiol and disulfide forms. A $E^{\circ\prime}$ value that is too low (that is, too negative) will destabilize the dithiol form and provide too little reactive thiolate to initiate attack of a nonnative disulfide bond. The CGPC active site of Trx, for example, has $E^{\circ\prime} = -0.27$ V (Krause *et al.*, 1991) and is an effective reducing agent. An enzyme that contains an active site with a $E^{\circ\prime}$ value that is too high will not be able to release trapped intermediates. The CPHC active site of DsbA has a $E^{\circ\prime} = -0.12$ V and is an effective oxidizing agent, but is inefficient at disulfide isomerization (Table 1.1) (Zapun *et al.*, 1993; Hennecke *et al.*, 1997; Huber-Wunderlich & Glockshuber, 1998).

The thiol pK_a , like the disulfide $E^{\circ\prime}$, must be in balance. An active-site cysteine residue not only acts as a nucleophilic thiolate to initiate catalysis, but also acts as a leaving group to allow for thiol–disulfide interchange (Gilbert, 1990). Hence, an isomerase will be optimal if its thiol pK_a equals the pH of the environment, which is near 7.0 in the ER (Hwang *et al.*, 1992). Although the pK_a of a typical cysteine residue is 8.7 (Szajewski & Whitesides, 1980), a thiol pK_a that is closer to 7.0 is more optimal for an oxidoreductase in the ER.

The *N*-terminal cysteine residue of the CXXC motif in thioredoxin-like enzymes consistently has a depressed pK_a value (6.7 for PDI; (Hawkins & Freedman, 1991)). This pK_a depression is consistent with the location of this residue at the *N*-terminus of an α-helix (Figure 1.2b; (Forman-Kay *et al.*, 1991)). The negative charge on the thiolate of the *N*-terminal cysteine residue is stabilized by interactions with the positive end of the α-helix dipole (Kortemme & Creighton, 1995). Changes in the intervening –XX– residues can enhance or diminish this Coulombic interaction and thereby decrease or increase, respectively, the pK_a of the *N*-terminal cysteine residue (Table 1.1) (Chivers *et al.*, 1996).

Variations in the –XX– residues of thioredoxin-like enzymes can also modulate the reduction potential of its active sites (Table 1.2). The variants can differ in thiol–disulfide interchange activity. Altering the residues in Trx can produce a more efficient oxidant (Joelson *et al.*, 1990; Krause *et al.*, 1991; Grauschopf *et al.*, 1995; Chivers *et al.*, 1996), while changing the residues in DsbA can make it a more potent reductant (Grauschopf *et al.*, 1995). The values of pK_a and $E^{\circ\prime}$ are related, as a more acidic thiol (lower pK_a) produces a less stable disulfide bond (higher $E^{\circ\prime}$) (Chivers *et al.*, 1997a). For –XX– variants of DsbA, changes in the disulfide $E^{\circ\prime}$ correlate well with changes in the pK_a of the *N*-terminal cysteine residue of the active site (Grauschopf *et al.*, 1995). On the other hand, active-site variants of Trx with different intervening residues reveal that thiol pK_a does not exclusively determine disulfide $E^{\circ\prime}$ (Table 1.2) (Chivers *et al.*, 1996; Chivers *et al.*, 1997a). Three active-site variants of Trx (CGHC Trx, CVWC Trx, and CWGC Trx) are able to catalyze the isomerization of disulfide bonds *in vivo* (Chivers *et al.*, 1996). For these variants, the changes in $E^{\circ\prime}$ do not correlate with changes in pK_a (Table 1.2), which is consistent with the presence of a pK_a-independent term in the relevant form of the Nernst equation (Chivers *et al.*, 1997a; Chivers & Raines, 1997):

$$E = E^{\circ} - \frac{RT}{nF}\ln\left(\frac{\alpha_0 F_{\mathrm{P(SH)_2}}}{[\mathrm{H}^+]^2 F_{\mathrm{PS_2}}}\right) \qquad \mathbf{1.1}$$

where E° is independent of pK_a and refers to the standard reduction potential, α_o refers to the fraction of fully protonated dithiol species, and $F_{\mathrm{P(SH)}_2}$ and F_{PS_2} are the formal (*i.e.*, total) concentrations for the reduced and oxidized molecules, respectively.

The effective concentration of two thiols reports on their tendency to form an intramolecular disulfide bond. In thioredoxin-like proteins, the effective concentration of thiols in the active site is determined largely by the three-dimensional structure of the protein (Burns & Whitesides, 1990; Chivers *et al.*, 1997a). The insertion of a tryptophan residue in CVWC Trx and CWGC Trx could destabilize the active-site disulfide bond simply by providing steric hindrance to its formation. Alternatively, removal of the proline residue in the CVWC and CGHC variants could increase the reduction potential by increasing the conformational entropy of the polypeptide chain, and thereby stabilizing the reduced form relative to the oxidized form. From such changes, a set of homologous oxidoreductases has evolved to carry out a variety of redox functions in different cellular environments (Table 1.1).

The Essential Function of PDI

PDI is required for the viability of *S. cerevisiae* (Farquhar *et al.*, 1991; LaMantia *et al.*, 1991; Scherens *et al.*, 1991; Tachikawa *et al.*, 1991). There has been some controversy over which of the cellular roles of PDI is necessary for survival. Initial work suggested that the most important function of PDI is the isomerization of nonnative disulfide bonds (Laboissière *et al.*, 1995b; Chivers *et al.*, 1998). Replacing PDI with a variant in which both active-site sequences are CGHS, instead of CGHC, restores viability to *pdi1*Δ *S. cerevisiae* (Table 1.3). Although this PDI variant has low dithiol oxidation activity and low disulfide reduction activity, it is proficient in its catalysis of disulfide isomerization (Table 1.3) (Laboissière *et al.*, 1995b; Walker *et al.*, 1996). A PDI variant that contains SGHC active-site sequences is unable to catalyze the formation, reduction, or isomerization of disulfide bonds, and does not complement a *pdi1* deletion (Table 1.3).

Further support for the conclusion that isomerization is the essential function comes from *in vivo* studies with PDI homologs. Overproduction of Eug1p (with its CLHS and CIHS active sites) is able to rescue *pdi1*Δ *S. cerevisiae*, even though the isomerization activity of Eug1p is less than that of PDI (Tachibana & Stevens, 1992). Trx is unable to restore viability to *pdi1*Δ yeast because of its low $E^{\circ\prime}$ value (Table 1.2) (Chivers *et al.*, 1996). A Trx variant in which the active site is replaced with CGPS is, however, able to complement a *pdi1* deletion (Table 1.2) (Chivers *et al.*, 1996). In Eug1p and CGPS Trx, catalysis of dithiol oxidation and disulfide reduction is not efficient because neither enzyme can form a disulfide bond in its active site. Nonetheless, both enzymes are able to catalyze disulfide isomerization and thereby endow a cell with the activity that is necessary for its survival.

These early studies suggest that only a small amount of isomerization activity is necessary for viability. The isolated a and a′ domains of rat PDI possess isomerization activity that is only 8.5% and 14%, respectively, that of wild-type PDI (Darby & Creighton, 1995b). These values agree well with the finding that having only a single reactive cysteine residue in the full-length protein results in a rate of sRNase A folding that is 10–12% lower than that of wild-type PDI (Walker & Gilbert, 1997). Even with this low activity, overexpression of the isolated a and a′ domains are able to rescue *pdi1*Δ *S. cerevisiae* with growth rates that are comparable to those with wild-type PDI (Xiao *et al.*, 2001). When overproduced, Eug1p is able to suppress a PDI deficiency in yeast, even though the *in vitro* isomerase activity of this CXXS enzyme is poor (Tachibana & Stevens, 1992; Nørgaard & Winther, 2001). Trx variants that complement a PDI deficiency are likely to be largely oxidized in the ER, and hence poor catalysts of isomerization (Table 1.2) (Chivers *et al.*, 1996; Chivers *et al.*, 1997b). These data may indicate that a minimal amount of reactive thiolate is required to initiate isomerization reactions in misfolded proteins in the ER.

An alternative explanation has been offered in recent work by Gilbert and coworkers (Xiao *et al.*, 2001; Solovyov *et al.*, 2004; Xiao *et al.*, 2004). Although the a′ domain of PDI possesses low isomerase activity (5% of wild-type yeast PDI), it is only 50% less efficient in oxidation. Using the inducible/repressible *GAL1–10* promotor systems to supplement a *pdi1*Δ *S. cerevisiae* strain with PDI variants, the effects of variable expression levels on growth and viability were examined. Supplementation with the a′ domain of yeast PDI was sufficient for viablity and required levels similar to the normal levels of PDI in wild-type *S. cerevisiae* (Solovyov *et al.*, 2004). These results suggest that either 60% of the oxidase activity or 5% of the isomerase activity of PDI is required for viability. The authors also found that viability with supplemented wild-type yeast PDI required approximately 60% of normal PDI levels (Solovyov *et al.*, 2004), suggesting that it is the oxidase activity of PDI that limits growth. Although both oxidization and isomerization are important in the cell, the authors conclude that disulfide isomerization may not be rate-limiting in the proteins essential for viability.

Balancing Isomerization and Oxidation

To catalyze disulfide isomerization, PDI must be in its reduced form. The ER, which has a $E_{\text{solution}} = -0.18$ V (Hwang *et al.*, 1992), seems to be the optimum environment for PDI to carry out oxidative protein folding. With $E^{\circ\prime} = -0.18$ V, PDI at equilibrium in the ER would exist as an equimolar mixture of reduced and oxidized forms (Lundström & Holmgren, 1993). Contrary to previous assumptions, studies with Ero1p revealed that the majority of PDI in the ER of *S. cerevisiae* exists in the oxidized state (Frand & Kaiser, 1999). This finding suggested that the majority of PDI is involved in disulfide formation rather than isomerization. A different result was obtained by modification of supplemented yeast PDI in *S. cerevisiae* with maleimide-conjugated polyethylene glycol which allowed more accurate gel-shift analysis. Using this method, it appears that PDI is at least partially reduced in the yeast ER with an average of 1.9 free active-site sulfhydryls available for initiation of isomerization (Xiao *et al.*, 2004).

The situation is somewhat different for human PDI. In three human cell lines (HeLa, COS, and U937) virtually all of the cellular PDI is reduced (Mezghrani *et al.*, 2001). The oxidized form of PDI only becomes detectable upon exposure of the cells to high levels of DTT. Perhaps, the redox system in human cells is more complex than is that in *S. cerevisiae* cells. For example, additional PDI and Ero homologs in human cells could allow PDI to be a specific catalyst of isomerization, as other enzymes take up the slack in oxidation.

Although the majority of PDI in *S. cerevisiae* may be devoted to catalyzing the formation of disulfide bonds, isomerization of disulfide bonds limits the rate of native disulfide formation in many cellular proteins. In *pdi1*Δ yeast, other proteins can fulfill the oxidative role of PDI, perhaps via the yeast UPR pathway (Laboissière *et al.*, 1995b; Chivers *et al.*, 1996; Nørgaard *et* *al.*, 2001; Xiao *et al.*, 2004), but to date no cellular redundancy has been discovered in the isomerization function of PDI.

1.4 Small Molecule Mimics of PDI

The primary determinants of isomerization efficiency are an $E^{\circ\prime}$ near –0.18 V and a thiol pK_a near 7.0. Using these criteria, it should be possible to design small-molecule mimics of Nature's enzymic isomerase. Such a small-molecule "foldase" would have a number of practical applications. Large-scale production of secretory proteins in *E. coli* is often complicated by aggregation or proteolysis associated with unfavorable conditions for disulfide formation in the bacterial cytosol (Marston, 1986; Gilbert, 1990). Insoluble protein aggregates must be solublized and folded *in vitro* in an appropriate redox buffer. Glutathione is often used for this task, but is not an efficient catalyst of oxidative protein folding. PDI is not a practical catalyst for large-scale protein production due to the high cost of its production, its instability, and the necessity to separate it from a target protein. An effective small-molecule catalyst would be desirable for oxidative protein folding *in vitro*, as well as for addition directly to cell cultures to improve heterologous protein production (Robinson *et al.*, 1994).

Although only one cysteine residue is required for catalysis of disulfide isomerization, the presence of two cysteine residues in the active site enhances the overall effectiveness of PDI in the formation of native disulfide bonds (Walker & Gilbert, 1997; Nørgaard & Winther, 2001). Hence, it is reasonable to assume that a small-molecule dithiol would be a more active catalyst of protein folding than would an analogous monothiol. This assumption has been confirmed by experiments. The dithiol (±)-*trans*-1,2-bis(mercaptoacetamido)cyclohexane (BMC or Vectrase™-P) has thiol pK_{a1} = 8.3 and pK_{a2} = 9.9 and $E^{\circ\prime}$ = –0.24 V (Table 1.4)

(Woycechowsky *et al.*, 1999). These values approach those of the CXXC motif in PDI, suggesting that this molecule may be efficient in catalyzing disulfide isomerization. Indeed, BMC increases both the rate of reactivation of sRNase A and the final yield of native RNase A over that attainable with glutathione (Woycechowsky *et al.*, 1999). In addition, BMC (a dithiol) shows an increase in activity and recovery over that with *N*-methylmercaptoacetamide (NMA, a monothiol) (Woycechowsky *et al.*, 1999). The presence of two sulfhydryl groups allows rescue from trapped mixed disulfides between catalyst and substrate, resulting in greater recovery of active RNase A and an increased rate of isomerization.

BMC is also an effective disulfide isomerase *in vivo*. *S. pombe* acid phosphatase is a 30-kDa homodimer with eight disulfide bonds (Robinson *et al.*, 1994; Wittrup, 1995). *S. cerevisiae* cells grown in the presence of BMC (0.1 g/L) secrete 3-fold more acid phosphatase, an increase equivalent to that achieved with 15-fold overproduction of PDI (Woycechowsky *et al.*, 1999). Likewise, the presence of BMC (2 μg/L) in an *E. coli* culture medium increases the yield of proinsulin, which has three disulfide bonds, by 60% (Winter *et al.*, 2002). These data indicate that a small-molecule mimic of PDI can enhance the yield of heterologous protein production in both eukaryotic and prokaryotic systems.

Aromatic thiols are also efficient small-molecule foldases. Aromatic thiols have low pK_a values (pK_a = 3–7) compared to their aliphatic counterparts (pK_a = 7–11) and are more reactive for thiol–disulfide interchange reactions at physiological pH (Whitesides *et al.*, 1977; Wilson *et al.*, 1977; Szajewski & Whitesides, 1980). Of particular interest is 4-mercaptobenzeneacetate, which was chosen because its pK_a value (pK_a = 6.6) closely resembles that of the active-site thiols of PDI (DeCollo & Lees, 2001). This reactive monothiol is able to catalyze the reactivation of sRNase A at a rate that is 5–6-fold higher than that of glutathione (Gough *et al.*, 2002). Further rate enhancement could be attainable with an aromatic dithiol.

Other design strategies for small-molecule foldases are based more literally on the active site of PDI. Linear CXXC peptides that mimic the active sites of a variety of oxidoreductases exhibit reduction potentials that do not reflect the enzyme from which they were derived. The reduction potential of the disulfide bond in the W–CGPC–KHI peptide, for example, is 70 mV lower than that of native CGPC Trx (Siedler *et al.*, 1993). Varying the –XX– sequence of linear CXXC octapeptides results in minimal changes in its redox properties with $E^{\circ\prime}$ values (calculated from equilibrium constants and a $E^{\circ\prime}$ of –0.252 for glutathione (Lees & Whitesides, 1993)) in the range of –0.220 V to –0.200 V (Siedler *et al.*, 1993).

On the other hand, by restricting the conformational freedom of the active-site motif in a cyclic hexapeptide, CXXC-containing peptides can be more efficient oxidative protein folding catalysts (Cabrele *et al.*, 2002). Three cyclic peptides, with sequences corresponding to glutaredoxin reductase (Grx), Trx, and PDI, show increasing activity and yield of native protein in a rRNase A assay. The enhancement in catalytic efficiency corresponds to a decrease in thiol pK_a and an increase in $E^{\circ\prime}$.

An alternative to changing the identity of intervening residues in a CXXC-containing peptide is to change the number of intervening residues in a $C(X)_nC$-containing peptide. To obtain an effective isomerase, it is desirable to have a fairly unstable disulfide bond. CXC peptides, with only one intervening residue, form 11-membered disulfide-bonded rings and typically have $E^{\circ\prime}$ values that are 30–40 mV lower than CXXC peptides (Kishore & Balaram, 1985; Zhang & Snyder, 1989; Moroder *et al.*, 1996). A CGC peptide is a promising candidate for a small-molecule isomerase (Woycechowsky & Raines, 2003). Although the pK_a of the reactive thiolate of CGC is higher than that of BMC, the reduction potential matches that of the PDI active site (Table 1.4). The CGC peptide is more efficient at folding sRNase A than is BMC. Even without imposing conformational constraints, a linear CGC is an effective catalyst of disulfide isomerization due to its favorable redox properties.

Small-molecule dithiols can act as isomerases, yet their activities lie far below that of PDI. Although the presence of a CXXC (or CXXS) motif is the only absolute requirement for isomerase activity, the protein scaffold provided by PDI increases the catalytic efficiency of the motif (Weissman & Kim, 1993). Not only does the protein scaffold form a conformationally constrained active site that serves to regulate its redox properties, it also provides a means of noncovalent interaction. In addition to covalent rearrangements, catalysis of disulfide isomerization relies on noncovalent interactions between PDI and its substrates. The b′ domain of PDI provides a binding site for unfolded proteins (Noiva *et al.*, 1993; Klappa *et al.*, 1998; Cheung & Churchich, 1999). Mutagenesis studies of individual PDI domains have shown that when combined with the a′ domain, the b′ domain improves the rate of BPTI isomerization by 75% over that with the a′ domain alone (Darby *et al.*, 1998). Further, full-length PDI is 7–12 fold more active in the folding of sRNase A than is the a or a′ domain alone (Darby & Creighton, 1995b).

Can an improved isomerase—one that combines the benefits of covalent and noncovalent interactions—be obtained by inserting a chemical catalyst into the context of a stable protein that has affinity for unfolded proteins? To answer this question, a CGC motif has been placed at the active site of Trx. Wild-type CGPC Trx has low activity in the catalysis of native disulfide formation due to the stability of its active-site disulfide bond (Pigiet & Schuster, 1986). The disulfide bond destabilization imposed by deletion of a proline residue in the active site endows reduced CGC Trx with the ability to reactivate sRNase A *in vitro* (Table 1.4; (Woycechowsky & Raines, 2003)).

The development of chemical catalysts that mimic the properties of PDI will be beneficial to both biotechnology and biomedicine. The production of eukaryotic proteins in *E. coli* often requires additional steps to obtain properly folded protein (Marston, 1986). Current techniques are neither time- nor cost-efficient. Small-molecule catalysts like BMC and CGC will provide better options for native disulfide formation during heterologous protein expression and purification from inclusion bodies. As a key player in the ER stress response (Dorner *et al.*, 1990), a small-molecule mimic of PDI could also be a useful chemotherapeutic in the treatment of protein folding diseases. Conversely, the catalytic activity of cell-surface PDI is necessary for the entry of HIV-1 virus into T lymphocytes (Gallina *et al.*, 2002; Matthias *et al.*, 2002; Barbouche *et al.*, 2003). Hence, cell-surface PDI is a valid target for inhibitor development.

1.5 Prolyl 4-Hydroxylase

Apart from its catalytic role in native disulfide bond formation, PDI plays an important role as a subunit of collagen prolyl 4-hydroxylase (P4H). In the ER lumen, P4H catalyzes the formation of (2*S*,4*R*)-4-hydroxyproline (Hyp) through the post-translational hydroxylation of (2*S*)-L-proline (Pro) residues in newly synthesized collagen strands. The enzyme is essential in *Caenorhabditis elegans* (Friedman *et al.*, 2000; Winter & Page, 2000), as the presence of Hyp is imperative for the formation of a stable collagen triple helix (Section 1.6).

P4H requires Fe(II), α-ketoglutarate, O_2, and ascorbate for hydroxylase activity (Prockop & Juva, 1965; Hutton *et al.*, 1966). During hydroxylation of proline, α-ketoglutarate is oxidatively decarboxylated to produce succinate and CO_2 (Figure 1.4; (Rhoads & Udenfriend, 1968)). One atom of O_2 is incorporated into the hydroxyl group of Hyp and the other into succinate (Cardinale *et al.*, 1971). Ascorbate is required to rescue inactivated enzyme that results from decarboxylation of α-ketoglutarate that is uncoupled from substrate hydroxylation (de Jong *et al.*, 1982).

Structural Properties of α Subunit of Prolyl 4-Hydroxylase

P4H is an $\alpha_2\beta_2$ tetramer (Berg & Prockop, 1973a) in which the α subunit (P4Hα) contains the enzyme active site (Helaakoski *et al.*, 1989) and the β subunit is PDI (Koivu *et al.*, 1987a; Pihlajaniemi *et al.*, 1987). The disulfide bond isomerization activity of PDI in the P4H complex is approximately half that of the free enzyme, due primarily to diminished activity of the *N*-terminal CGHC active site in the tetramer (Vuori *et al.*, 1992c). The reactive cysteine residues of PDI are not required for assembly of the tetramer or for hydroxylase activity, as SGHC active-site variants formed active tetramer in an insect cell expression system (Vuori *et al.*, 1992c). Instead, the functions of PDI in the complex are to retain P4H in the ER through the KDEL retention signal and to maintain the α subunit in a soluble, folded structure (Kivirikko *et al.*, 1989; Vuori *et al.*, 1992b; Vuori *et al.*, 1992c).

In the absence of PDI, the α subunit forms insoluble aggregates which cannot be refolded *in vitro* (Nietfeld *et al.*, 1981; Koivu & Myllylä, 1986). It is not surprising, therefore, that the *in vivo* assembly of α and β subunits occurs shortly after translation is completed or perhaps co-translationally (John *et al.*, 1993). Site-directed mutagenesis of the five conserved α subunit cysteine residues has identified two intramolecular disulfide bonds that are essential for P4H tetramer assembly (Lang & Scmid, 1988; John & Bulleid, 1994; Lamberg *et al.*, 1995). In the absence of sufficient oxidant, the α subunit aggregates and may associate with the ER-resident chaperone, BiP (John *et al.*, 1993; John & Bulleid, 1996). Heterologous production of P4H tetramer requires coexpression of the α and β subunits, and has been successful in insect and yeast cells (Vuori *et al.*, 1992b; Vuorela *et al.*, 1997).

Three isoforms of the P4Hα subunit, termed α(I), α(II), and α(III), have been identified in vertebrates (Helaakoski *et al.*, 1989; Helaakoski *et al.*, 1995; Annunen *et al.*, 1997; Kukkola *et al.*, 2003). The type I α subunit is the most prevalent and has been isolated from a variety of cell types. The type II enzyme comprises only 10–40% of the total P4H mRNA in most cell types (Annunen *et al.*, 1997) while the type III enzyme is present at even lower levels (Kukkola *et al.*, 2003). The human α(I) and α(II) subunits share 64% sequence identity while α(III) shares 35 and 37% identity with α(I) and α(II), respectively. All of the isoforms contain slightly more than 500 amino acid residues as well as a 17–21 residue signal sequence at the *N*-terminus. Most of the conserved amino acids occur in the *C*-terminal region where the catalytic residues have been identified.

Although no three-dimensional structure of P4H is available, site-directed mutagenesis studies have given some critical information about the amino acids located in the catalytic site. Mutation of the conserved residues His412 and His483 to serine and Asp414 to asparagine completely inactivate the enzyme but still assemble into tetramers and have wild-type values of K_m for a peptide substrate (Lamberg *et al.*, 1995; Myllyharju & Kivirikko, 1997). These three residues are likely ligands of iron in the P4H active site (Figure 1.5). Studies on a number of enzymes in the class of α-ketoglutarate-dependent Fe(II) oxygenases have revealed a common Fe(II) binding motif that includes two His residues and one Asp/Glu residue (Schofield & Zhang, 1999).

The α-ketoglutarate cosubstrate interacts with the P4H enzyme through three binding subsites which have been identified by studying the binding of a number of structural analogs of α-ketoglutarate (Majamaa *et al.*, 1984). Subsite I contains the positively charged Lys493 which binds to the C5 carboxyl group of α-ketoglutarate. Mutation of Lys493 to alanine or histidine completely inactivates the enzyme while a Lys493Arg variant retains 15% of wild-type activity (Myllyharju & Kivirikko, 1997). The K_m value for Fe(II), ascorbate and peptide substrate is unaffected by the latter mutation while the K_m value for binding to α-ketoglutarate is increased by 15-fold. In subsite II, the C1 carboxyl and C2 oxo functionalities of α-ketoglutarate are *cis*-coordinated to iron (Figure 1.5). Variants at His501 significantly impair enzyme activity by increasing the K_m for α-ketoglutarate (Lamberg *et al.*, 1995; Myllyharju & Kivirikko, 1997). Measurement of the binding constants for this variant to α-ketoglutarate analogs suggest that His501 helps to direct the binding of the C1 carboxyl group to Fe(II). Finally, subsite III is likely a hydrophobic pocket that accomodates the C3–C4 portion of the cosubstrate.

Catalytic Mechanism of Prolyl 4-Hydroxylase

Steady-state kinetic studies have demonstrated a sequential mechanism of catalysis for P4H (Myllyla *et al.*, 1977). Ferrous iron and α-ketoglutarate bind in an ordered fashion with the initial binding of iron (K_m = 5 μM) followed by the binding of α-ketoglutarate (K_m = 22 μM). There has been some disagreement over the order of binding of peptide substrate and O_2. While steady-state analysis of P4H suggests that O_2 (K_m = 40 μM) binds before the collagen mimic (Pro-Pro-Gly)$_{10}$ (Myllyla *et al.*, 1977), crystal structures of related enzymes in complex with Fe(II), α-ketoglutarate and substrate (Zhang *et al.*, 2000; Elkins *et al.*, 2002; Müller *et al.*, 2004), as well as spectroscopic and molecular modeling studies (Zhou *et al.*, 1998; Ryle *et al.*, 1999; Ho *et al.*, 2001), predict the opposite. It appears likely that peptide substrate binds in close proximity to iron which causes a conformational shift in the iron active site, releasing a water ligand and allowing an open site for O_2 binding.

A putative mechanism for proline hydroxylation by P4H is detailed in Figure 1.5. Although none of the proposed intermediates have been directly observed for P4H, evidence for this mechanism has come from the study of a number of related enzymes in the large superfamily of α-ketoglutarate-dependent non-heme iron(II) oxygenases, including clavaminate synthase (CS), taurine:α-ketoglutarate dioxygenase (TauD), and thymine hydroxylase among others (reviewed in (Costas *et al.*, 2004)). Broadly, the reaction involves the formation of a reactive hydroxylating species followed by the hydroxylation of peptide substrate.

Oxygen binding to the open site of Fe(II) leads to the formation of an Fe(II)-superoxide radical anion species. Evidence for this intermediate is provided by the observation that superoxide scavengers are competitive inhibitors of P4H with respect to O_2 (Tuderman *et al.*, 1977; Myllyla *et al.*, 1979) and from DFT calculations on the CS enzyme (Borowski *et al.*, 2004). Nucleophilic attack of superoxide on the α-keto group of the cosubstrate may form an Fe(IV) peroxo bicyclic species as the first step towards the formation of the activated oxygen species. DFT calculations on CS indicate that this species is spontaneously converted to an

Fe(II)-peracid species by cleavage of the C–C bind to form the first product, CO_2 (Borowski *et al.*, 2004). Two electrons are then donated from iron to produce an Fe(IV)=O species. Although evidence for the Fe(IV) species remains to be obtained for P4H, recent studies with the related enzyme, TauD, have provided extensive spectroscopic evidence for such a species. Rapid-freeze quench Mössbauer, EPR, and x-ray absorption spectroscopy have identified a high-spin Fe(IV) species (Price *et al.*, 2003b; Riggs-Gelasco *et al.*, 2004). Further, resonance Raman spectroscopy under cryogenic conditions has detected an Fe(IV)=O species (Proshlyakov *et al.*, 2004).

The reactive iron-oxo intermediate catalyzes prolyl hydroxylation in a two-step process involving radical formation followed by oxygen rebound (Groves & McClusky, 1976). Kinetic isotope effects observed for P4H (Wu *et al.*, 2000), thymine hydroxylase (Thornburg *et al.*, 1993) and CS (Iwata-Reuyl *et al.*, 1999) support the two-step mechanism in which hydrogen atom abstraction is the rate-limiting step to hydroxylation. A radical probe substrate has also been used to suggest a prolyl radical intermediate in the P4H mechanism (Wu *et al.*, 1999). Support for the hypothesis that the Fe(IV) species is the hydroxylating intermediate comes from the observation that substitution of the C1 hydrogen of taurine with deuterium slows the decay of the TauD–Fe(IV) intermediate by 37-fold, as determined by stopped-flow absorption and freeze-quench Mössbauer spectroscopy (Price *et al.*, 2003a).

In addition to catalyzing the hydroxylation of prolyl residues, the P4H enzyme can also catalyze the decarboxylation of α-ketoglutarate in the absence of peptide substrate (Counts *et al.*, 1978). This uncoupled decarboxylation is also observed to a small degree in the presence of peptide substrate and leads to inactivation of the enzyme (Myllylä *et al.*, 1978). In the absence of peptide substrate, ascorbate consumption is stoichiometrically coupled to CO_2 production, whereas in the presence of peptide substrate, ascorbate is consumed at a much lower rate

(Myllylä *et al.*, 1984). Although ascorbate is not absolutely required in the initial cycles of catalysis, ascorbate is necessary to protect P4H against inactivation that occurs as a result of the uncoupled reaction (Myllylä *et al.*, 1978). In the absence of ascorbate, the accumulation of Fe(III) and a radical species have been observed by EPR (de Jong *et al.*, 1982). Inactivation of P4H likely occurs by conversion of the reactive iron-oxo species to Fe(III) and an OH radical which can attack enzyme residues near the active site (Nietfeld & Kemp, 1981; de Jong *et al.*, 1982). Ascorbate (K_m = 330 μM) binds in a similar site as α-ketoglutarate (Majamaa *et al.*, 1986) to reduce the Fe(III) species before it can do irreversible damage to the enzyme.

Two enzymes of the α-ketoglutarate-dependent Fe(II) oxygenase family have identified specific residues affected by the uncoupled decarboxylation reaction. Reaction of TfdA (2,4-dichlorophenoxyacetate/α-ketoglutarate dioxygenase) with O_2 in the presence of α-ketoglutarate and Fe(II) leads to hydroxylation of Trp112 and the development of an Fe(III)–O–Trp chromophore (Liu *et al.*, 2001). Resonance Raman spectroscopy identified a catecholate species under similar conditions with TauD. A combination of stopped-flow spectroscopy, EPR spectroscopy, and mutational studies identified a Tyr73 radical that inactivates the enzyme in the absence of ascorbate (Ryle *et al.*, 2003). Similar studies have yet to be performed with P4H.

Substrate Recognition by Prolyl 4-Hydroxylase

Collagen strands are characterized by repeating Xaa–Yaa–Gly units, in which Xaa and Yaa can be any amino acid, though Pro is often found in the Xaa position and Hyp is often found in the Yaa position (Section 1.6). P4H catalyzes hydroxylation of Pro residues almost exclusively in the Yaa position of collagen strands (Hutton *et al.*, 1967). The hydroxylation reaction is performed on individual collagen chains and is prevented by the triple helical conformation

(Berg & Prockop, 1973b). The minimum substrate required for hydroxylation is an Xaa–Pro–Gly triplet, but the K_m for Pro–Pro–Gly binding to the α(I) tetramer is high (20 mM) (Prockop *et al.*, 1976). Although Pro is the preferred residue in the Xaa position, hydroxylation can occur with substitution of a variety of amino acids at this position, albeit at lower reaction rates (Kivirikko *et al.*, 1972; Rapaka *et al.*, 1978).

Peptides of increasing chain length have progressively lower K_m values while values for V_{max} do not change significantly (Table 1.5). For example, the K_m for (Pro–Pro–Gly)$_5$ is 10-fold higher than the K_m for (Pro–Pro–Gly)$_{10}$ when expressed in terms of Pro–Pro–Gly units (Myllyla *et al.*, 1977; Vuori *et al.*, 1992b; Hieta *et al.*, 2003). A type I protocollagen strand (MW = 150,000) has an even tighter binding constant (K_m = 0.2 μM) (Annunen *et al.*, 1997). The effect of chain length on binding constants for P4H cannot be explained by cooperation between the two α-subunit active sites or substrate binding sites as the same binding trends are observed for an αβ dimer as for an $\alpha_2\beta_2$ tetramer (Kukkola *et al.*, 2004). The K_m values for the α(II) tetramer are consistently 3–6 times higher than those for the α(I) tetramer (Annunen *et al.*, 1997; Hieta *et al.*, 2003), while the α(III) tetramer has substrate K_m values intermediate to α(I) and α(II) (Kukkola *et al.*, 2003).

A region of approximately 100 residues (Gly138–Ser244) has been identified, through limited proteolysis studies, as the peptide binding domain of P4H (Myllyharju & Kivirikko, 1999). This region, which is separate from the catalytic domain, contains a large number of hydrophobic and aromatic residues. Ile182 and Tyr233 in the α(I) subunit (Glu and Gln, respectively, in the α(II) subunit) account for much of the difference in substrate binding between the α(I) and α(II) enzymes (Myllyharju & Kivirikko, 1999). The substrate domain was expressed in *E. coli* and its binding properties for synthetic peptide substrates measured. The K_d values measured for a range of substrates are nearly identical to the K_m values for binding to the full-length P4H tetramer (Hieta *et al.*, 2003).

Recently, crystallographic study of the this region (Phe144–Ser244) has allowed our first glimpse of the P4H structure (Pekkala *et al.*, 2004). The peptide substrate binding domain contains five α helices which are organized to form a "bowl-like surface". The structure can be divided into two regions each with high structural similarity to tetratricopeptide repeat (TPR) domains which are typically involved in protein-protein interactions (D'Andrea & Regan, 2003). The two TPR regions are linked by a relatively long loop which is fairly conserved in various P4Hα subunits. The concave surface of this domain contains a patch of hydrophobic amino acids that likely compose a binding groove for peptide substrate. In particular, three tyrosine residues are spaced at approximately 8.5 Å intervals, similar to the repeat distance of the –Pro–Xaa–Xaa–Pro– motif of a collagen substrate with a polyproline type-II (PP-II) helix.

A unique conformational model for the binding of peptide substrate to P4H has been proposed (Atreya & Ananthanarayanan, 1991). From studies of the conformational features of synthetic peptide substrates and their binding properties, it has been suggested that P4H substrates bind using both a PP-II helix and a β-turn. The β-turn structure may be the minimum requirement for catalysis in the active site while the PP-II helix enhances binding by interaction with substrate binding subsites (Figure 1.6).

1.6 Collagen

Collagen is the most abundant protein in animals. In humans, this large family of extracellular matrix proteins includes 27 types of collagen which are important for the integrity and strength of bone, skin, cartilage, and tendon (Myllyharju & Kivirikko, 2004; Yang *et al.*, 2004). The unique tertiary structure of collagen is formed from three parallel, left-handed, polyproline II strands which are wrapped around each other into a triple helix with a shallow, right-handed superhelical pitch (Figure 1.7a). The majority of triple helical collagens further organize into fibrils which provide both strength and flexibility to their constituent tissues.

Crystallographic studies of collagen mimics have confirmed early structural work based on x-ray diffraction analyses of biological samples (Rich & Crick, 1961; Bella *et al.*, 1994; Kramer *et al.*, 1999; Nagarajan *et al.*, 1999; Kramer, 1998 #314; Berisio *et al.*, 2001). The triple helix, composed of repeating Xaa–Yaa–Gly tripeptides, is stabilized in part by the hydrogen bonding interactions between the NH of Gly and the C=O of the Xaa residue (Figure 1.7b and c). The superhelical twist appears to be somewhat sequence dependent as imino acid rich regions have a 7/2 helical symmetry (3.5 residues per turn) while imino acid poor regions have a 10/3 helical twist (3.3 residues per turn) (Kramer *et al.*, 1999). The placement of Gly at every third residue is required by the triple helical structure since it is buried in the tightly packed interior of the triple helix. Residues in the Xaa position are exposed to the solvent while Yaa position residues are less solvent-accessible due to their interaction with neighboring chains.

Hydroxyproline and Collagen Stability

The Xaa and Yaa residues of the repeating Xaa–Yaa–Gly tripeptides can be any amino acid. Analysis (Ramshaw *et al.*, 1998) of the natural occurence in various collagens, however, reveals that the Xaa position is most commonly occupied by Pro and the Yaa residue is most often Hyp. The only triplet that occurs with high frequency is Pro–Hyp–Gly (10.5%). The presence of Hyp in the Yaa position stablizes the triple helix of native collagen as well as peptide mimics of collagen (Berg & Prockop, 1973c; Sakakibara *et al.*, 1973; Holmgren *et al.*, 1998; Ramshaw *et al.*, 1998). Pro hydroxylation is required for collagen stability at physiological temperatures—a fully hydroxylated type I collagen has a melting temperature (T_m, the temperature at the midpoint of the thermal transition) of 43 °C, while an unhydroxylated form has a T_m of only 27 °C (Berg & Prockop, 1973c). The basis for this enhanced stability has been the subject of extensive research (reviewed in (Jenkins & Raines, 2002)).

The preferred hypothesis for many years was that hydrogen bonding interactions mediated by an extensive network of water molecules, could explain the extra stability provided by Hyp residues. X-ray crystallographic studies suggested that the triple helix is highly hydrated and that the hydroxyl group of Hyp binds to a main-chain carbonyl of a neighboring strand through one or two water molecules (Bella *et al.*, 1994; Kramer *et al.*, 1998). Contradictory data were offered from the crystal structures of $(\text{Pro–Pro–Gly})_{10}$ and $(\text{Pro–Hyp–Gly})_{10}$ obtained by Okuyama and colleagues which revealed a less substantial hydration network (Nagarajan *et al.*, 1998; Nagarajan *et al.*, 1999). In addition, the same number of water molecules were observed for both structures, suggesting that the stability of the Hyp-containing helix was not due to water-mediated hydrogen bonds. This conclusion was supported by the finding that Hyp confers stability to the triple helix even in anhydrous conditions (Engel *et al.*, 1977).

An alternative explanation has been provided by studies with Hyp analogs in which the hydroxyl group is replaced with a fluorine atom. Collagen mimic peptides which contain 4(*R*)-fluoro-L-proline (Flp) in the Yaa position, $(\text{Pro–Flp–Gly})_{10}$, assemble into triple helices. The T_m of these helices is 91 °C in 50 mM acetic acid, indicating increased stability relative to $(\text{Pro–Hyp–Gly})_{10}$ with a T_m of 69 °C (Holmgren *et al.*, 1998). Organic fluorine has a very low tendency to form hydrogen bonds (Holmgren *et al.*, 1999), thus the stabilization of Flp in the

Yaa position cannot be attributed to hydrogen bonding. Instead, the stabilization confered by Flp (and Hyp) has been explained by inductive effects generated from the electron-withdrawing substituent on the proline ring (Holmgren *et al.*, 1998; Holmgren *et al.*, 1999; Bretscher *et al.*, 2001; Jenkins & Raines, 2002).

The pyrrolidine ring pucker of Hyp residues predominantly occupies the C^{γ}-*exo* (up) conformation due to the inductive effect of the hydroxyl group (Bretscher *et al.*, 2001). Analysis of the crystal structure of (Pro–Pro–Gly)$_{10}$ revealed that residues in the Yaa position prefer torsion angles [$\varphi = (-60.1 \pm 3.6)^\circ$ and $\psi = (152.4 \pm 2.6)^\circ$] that correspond to the C^{γ}-*exo* conformation (Berisio *et al.*, 2002). Residues in the Xaa position, however, prefer torsion angles [$\varphi = (-74.5 \pm 2.9)^\circ$ and $\psi = (164.3 \pm 4.3)^\circ$] that correspond to the C^{γ}-*endo* (down) conformation (Berisio *et al.*, 2002). The electronegativity of the hydroxyl group also increases the *trans*:*cis* ratio of the amide bond, stabilizing the *trans* conformation of Hyp, a prerequisite for the collagen triple helix (Bretscher *et al.*, 2001). Finally, the electron-withdrawing substituent preorganizes the ψ angle of the peptide bond (141° for crystalline AcFlp–OMe) (Panasik *et al.*, 1994; Bretscher *et al.*, 2001). The stereoelectronic preorganization of φ, ψ, and ω dihedral angles thus stabilizes the collagen triple helix when Hyp residues are located in the Yaa position. This analysis also explains the observation that Hyp in the Xaa position destabilizes the triple helix (Inouye *et al.*, 1982), as the dihedral angles enforced by the C^{γ}-*endo* ring pucker would introduce conformational strain (Bretscher *et al.*, 2001; Hodges & Raines, 2003).

Triple Helix Assembly in Fibrillar Collagens

The assembly of fibrillar collagens into a triple helix requires a series of post-translational modifications. Individual strands are translated in a precursor form, termed procollagen, and are directed to the ER by an *N*-terminal leader sequence. Upon insertion into the lumen, P4H catalyzes the hydroxylation of specific proline residues (Section 1.5), and lysine hydroxylase hydroxylates a small subset of Lys residues in the Yaa position (Kivirikko & Myllyla, 1980; Kivirikko & Pihlajaniemi, 1998). Hydroxylysine (Hyl) residues are further modified by the addition of a galactose monosaccharide, or a glucose-galactose disaccharide (Kivirikko & Myllylä, 1982). Once the peptide is fully synthesized and hydroxylated, triple helix assembly can begin.

Each procollagen chain contains *C*- and *N*-terminal propeptides which assist in the association of individual strands (Fessler *et al.*, 1975). Formation of intramolecular disulfide bonds in the *N*- and *C*-propeptides is likely catalyzed by PDI (Lees & Bulleid, 1994). Following oxidation, individual chains must associate with each other. Different collagen types may form homo- or hetero-trimers. Type I collagen, for example, is composed of two proα1(I) chains and one proα2(I) chain, while Type III collagen is a homotrimer of three proα1(III) chains. Each cell may synthesize as many as six different chains at the same time which must associate in the appropriate manner (Byers, 2001). An initial recognition event, directed by a 15-residue domain in the *C*-propeptide, is responsible for ensuring selective chain association (Lees *et al.*, 1997). Although not required, interchain disulfide bonds further stabilize the association of the three chains and ensures the correct register for the triple helix (Bulleid *et al.*, 1996; Lees *et al.*, 1997). Triple helix formation begins with a nucleation event at the *C*-terminus of the triple helix domain which requires the presence of 5–6 Hyp residues (Bulleid *et al.*, 1996). The triple helix is then propagated in the *C*- to *N*-terminal direction in a zipper-like manner (Engel & Prockop, 1991). This process is rate-limited by the *cis-trans* isomerization of the imino acid peptide bonds (Roth & Heidemann, 1980; Bächinger, 1987).

Upon triple helix formation, mature collagen is secreted and the propeptides are released by the action of specific proteases (Prockop *et al.*, 1998). Propeptide cleavage decreases the solubility of the collagen molecule and induces the assembly of collagen into fibrils (Kadler *et al.*, 1987). The fibrillar structure varies depending on collagen type and tissue distribution, but is arranged through the head-to-tail lateral alignment of collagen molecules. Proper alignment leads to crosslinking of the collagen chains (reviewed in (Eyre *et al.*, 1984)). Lysyl oxidase converts specific Lys and Hyl residues to aldehydes which react with nearby Lys, Hyl and His residues to form intermolecular crosslinks which stabilize the fibrillar structure.

Collagen-Related Diseases

Collagen mutations are responsible for a large number of hereditary connective tissue diseases. The most common mutations in the fibrillar collagens are Gly substitutions (Kuivaniemi *et al.*, 1997). Osteogenesis imperfecta or "brittle bone" disease can result from more than 150 different Gly substitutions in Type I collagen (Byers, 1993). Various phenotypes are observed, from mild to lethal, and depend on the distance of the mutation from the *C*-terminal nucleation site and the local environment of the mutation. Mutation of Gly to serine, cysteine, arginine, or aspartate results in a slow rate of triple helix formation as indicated by an increase in Hyl formation (Kirsch *et al.*, 1981). Although folding is often achieved, the mutant triple helices are characterized by decreased thermal stability and local conformational perturbations which can cause abnormal fibril formation. Similar mutations in the cartilagenous Type II collagen lead to chondrodysplasia (Prockop & Kivirikko, 1995). Mutations can also occur in the Xaa and Yaa positions. Early onset osteoarthritis, for example, is caused by a Type II collagen mutation of arginine to cysteine in the Yaa position (Ala-Kokko *et al.*, 1990).

Ehlers-Danlos syndrome (EDS) is characterized by joint hypermobility, excessive skin elasticity, abnormal bruising and skeletal deformities (reviewed in (Mao & Bristow, 2001)). Various forms of the disease are caused by defective RNA splicing events, gene insertions, deletions, or rearrangements, or missense mutations to the collagen genes. In some cases, these mutations may alter the rate of fibril formation or the fibril structure. Type VII EDS is caused by point mutations in Type I collagen that affect proteolytic processing of the *N*-propeptide. Type VI EDS is the result of a lysyl hydroxylase deficiency.

The fibrotic diseases, including liver cirrhosis, rheumatoid arthritis, pulmonary fibrosis, and scleroderma, are caused by the progressive accumulation of connective tissue, primarily composed of collagen (reviewed in (Trojanowska *et al.*, 1998)). Pathological fibrosis is most commonly observed in the liver, lungs, kidney, and skin as a result of an abnormal response to tissue damage. Injury to tissues, as a result of infection, diabetes, hypertension, toxic chemicals or trauma, activates an inflammatory response which leads to an increase in collagen production through the activation of fibroblasts and fibroblast-like cells. Fibrosis results when the termination stage of normal wound healing becomes unregulated. Ultimately, the excessive accumulation of extracellular matrix will lead to a loss of tissue function.

1.7 A Need for More Mechanistic Understanding

What are the precise chemical mechanisms by which PDI achieves disulfide formation and isomerization? How does P4H catalyze hydroxylation of Pro residues and stabilize the collagen triple helix? What is the substrate specificity of PDI, and its molecular basis? How is PDI able to balance its role as a disulfide isomerase and dithiol oxidase? How does P4H recognize specific Pro residues along a collagen strand for hydroxylation? Major obstacles hinder researchers seeking answers to these and related questions. The lack of a three-dimensional structure for intact PDI or for tetrameric P4H makes it difficult to obtain a clear picture of the active-site residues within each enzyme. Researchers are also lacking a clear picture of how the enzymes interact with their substrates. Further, substrates now available for studying disulfide isomerization are too complex to delineate the details of the chemical mechanism of PDI. The methods available for assaying catalytic activity of P4H are inefficient and imprecise, making it difficult to perform rigorous mechanistic studies.

Future Directions for Protein Disulfide Isomerase

The structures of individual a and b domains of PDI are known (Figure 1.1b; (Kemmink *et al.*, 1996; Kemmink *et al.*, 1997)) and provide evidence for structural similarity between all four of the major domains. It is not yet understood either how the four domains interact with each other or how the two active sites communicate with each other. At saturating concentrations of substrate, the two active sites play different roles. The *N*-terminal active site is responsible for most of the catalytic activity, while the *C*-terminal active site contributes more to substrate binding (Lyles & Gilbert, 1994). At substrate concentrations near the value of K_M (where the rate is half-maximal) the two active sites are equally efficient catalysts and function independently (Walker *et al.*, 1996). In contrast, communication between the two active sites plays an integral role in catalysis by *E. coli* DsbB (Kadokura & Beckwith, 2002) and *S. cerevisiae* Ero1p (Frand & Kaiser, 2000).

Structures of PDI homologs provide some intriguing proposals for the arrangement of individual PDI domains. CVWC Trx crystallizes in a dimer-like structure (Schultz *et al.*, 1999). This dimer is connected through a continuous β-sheet across both monomers and interlocked α-helices that cap the β-strands. Similarly, the structure of a disulfide oxidoreductase from *Pyrococcus furiosus* contains two CXXC-containing thioredoxin-like units that are joined by a continuous β-sheet (Ren *et al.*, 1998). In both of these examples, adjacent thioredoxin-like units are arranged colinearly.

The structure of DsbC provides useful information on that of PDI (McCarthy *et al.*, 2000). DsbC catalyzes disulfide isomerization in the bacterial periplasm (Zapun *et al.*, 1995; Reitch *et al.*, 1996). DsbC is a dimeric protein, and, like the enzymes described above, its monomers are joined through a continuous β-sheet. Each monomer is composed of a catalytic *C*-terminal domain and an *N*-terminal dimerization domain. This arrangement is similar to the arrangement of domains in PDI in which the a and a′ domains are catalytic and the b and b′ domains serve another role. In the DsbC crystal structure (McCarthy *et al.*, 2000), the *N*- and *C*-terminal domains are connected by a hinged α-helix that allows for flexibility of the catalytic domain. Between the two active sites is a broad uncharged cleft that can accommodate the nonspecific binding of misfolded substrate proteins. The flexibility provided by the hinged region allows for variation in the size of the cleft to allow binding of large or small substrate proteins. This structure provides a model for how the domains of PDI might interact to allow the nonspecific binding of substrates.

A second bacterial protein, DsbG, also catalyzes disulfide bond isomerization in the bacterial periplasm. DsbG shares 24% sequence identity with DsbC, but is expressed at lower levels and has a more narrow substrate specificity (Bessette *et al.*, 1999b). Crystallization studies of DsbG reveal a similar structure as DsbC in which a V-shaped dimer is formed through *N*-terminal association domains (Heras *et al.*, 2004). The cleft of the DsbG dimer, however, is almost twice the size of that of DsbC, and is lined with charged residues. These structural differences suggest how DsbG is able to recognize specific substrates which likely bind as partially folded proteins (Heras *et al.*, 2004).

The structure and function of DsbC in a complex with the *N*-terminal domain of DsbD (DsbDα) provides insight on how a disulfide isomerase interacts with a protein substrate (Goldstone *et al.*, 2001; Haebel *et al.*, 2002). DsbDα binds in the cleft of DsbC, causing conformational changes in both the substrate and catalyst. Both active sites of DsbC contact DsbDα cysteine residues. The primary noncovalent interactions are between hydrophobic or uncharged polar groups.

Even with more structural information, a proper disulfide isomerization substrate for detailed mechanistic work is essential. Currently, the most common *in vitro* substrates for studying catalysis by PDI are RNase A and BPTI, neither of which is ideal. RNase A is a 124-residue protein with 8 cysteine residues that can be randomly oxidized under denaturing conditions to give sRNase A, a mixture of up to 105 different fully oxidized conformations, only one of which is native (Konishi *et al.*, 1981; Scheraga *et al.*, 2001). BPTI, a 58-residue protein, is a less complex substrate with only 6 cysteine residues. Still, BPTI can form 15 different fully oxidized species on the pathway to its native structure (Weissman & Kim, 1993).

Simple peptide substrates have been used to study disulfide formation and reduction by PDI. A 28-residue peptide, based on the sequence of BPTI, with only one disulfide bond was an effective substrate for oxidation by PDI in the presence of glutathione (Darby *et al.*, 1994). A fluorescently-labeled substrate of seven amino acids and one disulfide bond has been used to study the oxidation and reduction activities of PDI (Westphal *et al.*, 1998). Recently, a redox sensitive variant of green fluorescent protein (GFP), with an additional disulfide bond, has been developed as a tool to monitor disulfide bond formation in living cells (Østergaard *et al.*, 2001).

Likewise, a simple substrate could be used to characterize disulfide isomerization by PDI in an *in vitro* assay. The ideal substrate would contain only two disulfide bonds and thus only three fully oxidized forms. To be a substrate for characterizing disulfide isomerization activity, the two disulfide bonds in the native form must be both stable under the conditions of a typical folding assay ($E_{\text{solution}} = -0.18$ V) and more stable (lower $E^{\circ\prime}$) than those in the other two fully oxidized forms. A continuous folding assay could be used with such a substrate. For example, by modifiying the substrate to contain an appropriate donor and acceptor, fluorescence resonance energy transfer (FRET) could be used to monitor any conformational change that accompanies disulfide isomerization. Unlike the discontinuous assays currently available, a continuous folding assay would permit rapid, detailed analysis of the chemical and kinetic mechanism by which PDI catalyzes the isomerization of nonnative disulfide bonds.

Future Directions for Research on Prolyl 4-Hydroxylase

As with PDI, the three-dimensional structure of P4H tetramer has not yet been determined. Crystallographic data from a number of enzymes in the α-ketoglutarate-dependent Fe(II) oxygenase family have given some insight into the structure of the catalytic site of P4H. The structure of deacetoxycephalosporin C synthase (DAOCS) was the first of this enzyme family to be solved (Valegård *et al.*, 1998), and revealed that the ferrous active site is located in a distorted jelly roll motif. Subsequent crystal structures of CS (Zhang *et al.*, 2000), proline 3-hydroxylase

(Clifton *et al.*, 2001), anthocyanidin synthase (ANS; (Wilmouth *et al.*, 2002), TauD (Elkins *et al.*, 2002), an alkylsulfatase (AtsK; (Müller *et al.*, 2004), and factor-inhibiting hypoxia-inducible factor 1 (FIH-1; (Dann *et al.*, 2002; Elkins *et al.*, 2003) have confirmed a common jelly roll motif for this class of enzymes even though they show a very low degree of sequence homology. Strikingly, their active sites, containing the common 2 His/1 carboxylate motif, are all accomodated in equivalent positions within this fold. It seems likely that P4H contains its active site within a similar jelly roll motif, although the precise residues involved have not been determined.

The structural differences between family members occur in domains surrounding the jelly roll core which are primarily involved in substrate binding. DAOCS, proline 3-hydroxylase, ANS, and FIH-1 (Valegård *et al.*, 1998; Clifton *et al.*, 2001; Dann *et al.*, 2002; Wilmouth *et al.*, 2002) all have a *C*-terminal α-helical bundle that appears to be flexible and may serve a roll in substrate binding. CS, AtsK, and TauD, on the other hand, contain a domain, combining α helices and β strands, that is inserted within the jelly roll motif (Zhang *et al.*, 2000; Elkins *et al.*, 2002; Müller *et al.*, 2004). A common feature for all of these proteins is at least one flexible region near the active site that may be important for regulating the binding of substrate.

The multimeric structure of some of these proteins may provide clues to the organization of subunits in the P4H tetramer. The dimer interface of FIH-1 is a helical bundle composed of helices from the *C*-terminus of each monomer (Dann *et al.*, 2002; Elkins *et al.*, 2003; Lee *et al.*, 2003). Dimers of AtsK are formed through interaction of α-helical residues in the insert region (Müller *et al.*, 2004). Dimers of AtsK interact to form tetramers through hydrophobic interactions between β-sheets of the jelly roll motif (Müller *et al.*, 2004). This tetrameric structure reveals a central cavity which is not entirely isolated from solvent. It is not yet understood how the two α subunits of P4H interact with each other or with PDI.

The members of the α-ketoglutarate-dependent Fe(II) oxygenase family bind a wide variety of substrates. The recent crystal structure of FIH-1, an aspariginyl hydroxylase, with peptide substrates of 40–50 residues (Elkins *et al.*, 2003) suggest a mode of binding for an extended polypeptide chain to this enzyme class. The structure reveals two sites of interaction between peptide and enzyme. In the first site, the peptide binds through hydrogen bond interactions to a groove that extends from the surface into the buried active site. The substrate Asn residue is involved in a tight turn and precisely oriented in the active site by three hydrogen bonds which direct the side chain towards the Fe(II). In the second binding site, the more *C*-terminal peptide residues form an α helix that binds in a hydrophobic pocket composed of several Leu residues. This organization seems analogous to the proposed mode of binding of a procollagen substrate to P4H (Atreya & Ananthanarayanan, 1991), in which an extended PP-II helix binds in a hydrophobic groove (Pekkala *et al.*, 2004) while the Pro to be hydroxylated is involved in a β turn at the active site. The recent crystal structure of the substrate binding domain (Phe144–Ser244) of human P4H α(I) has confirmed that a patch of hydrophobic residues are involved in peptide binding and appears ideal for binding of a collagen PP-II helix (Pekkala *et al.*, 2004). IndeFurther structural insights which could guide future enzymological studies will require a crystal structure of the complete P4H tetramer.

The design of future enzymological studies is also hindered by the available assays for measuring prolyl hydroxylation. The traditional assay for P4H activity measures hydroxylation of (Pro–Pro–Gly)$_n$ collagen mimics indirectly—by monitoring the release of [^{14}C]CO$_2$ during decarboxylation of [1-^{14}C]α-ketoglutarate (Rhoads & Udenfriend, 1968; Kivirikko & Myllylä, 1982). CO_2(g) release necessarily provides an indirect measurement of hydroxylation, which can be problematic. In the absence of sufficient substrate, P4H mediates the uncoupled decarboxylation of α-ketoglutarate, which does not produce Hyp and inactivates the enzyme by an as yet unidentified mechanism (Myllylä *et al.*, 1984; Wu *et al.*, 2000). Even under conditions of saturating substrate, this uncoupled reaction occurs to a small degree. For this reason, ascorbate is included in all activity assays to rescue enzyme trapped in the Fe(III) state. In addition, the CO_2-release assay method is inefficient. First, the complexities of trapping a gaseous product are difficult to avoid. Second, the assay requires comparatively large amounts of enzyme, as the measurement of [^{14}C]CO_2 release requires that each time point derive from a separate reaction mixture.

Other less common assays that monitor hydroxylation directly use radiolabeled procollagen substrates (Juva & Prockop, 1966; Kivirikko & Myllylä, 1982). These assays have additional disadvantages. They require the tedious separation of ^{14}C- or ^{3}H-labeled hydroxyproline from radiolabeled proline prior to scintillation counting. An assay that allows for the direct analysis of a hydroxylation reaction mixture in an automated manner is desirable. The study of the enzymatic mechanism of hydroxylation and the development of inhibitors of the fibrotic diseases would benefit from the development of a direct and efficient assay of prolyl hydroxylation.

This thesis seeks to improve the available tools for answering some of the intriguing questions that remain about the mechanism of action of PDI and P4H. The development of a continuous assay for disulfide bond isomerization has allowed the rapid characterization of active-site variants of PDI (Chapter 2). The method described will permit future studies into the mechanism by which PDI catalyzes isomerization and characterization of small-molecule isomerases. An *E. coli* expression system for the P4H tetramer and the development of a direct, efficient HPLC-based assay for prolyl hydroxylation will facilitate future mechanistic studies of the enzyme (Chapter 3). These studies will include the development of P4H inhibitors for therapeutic applications as well as understanding the mechanism by which Hyp stabilizes the collagen triple helix (Chapter 4).

Table 1.1 Properties of oxidoreductases

Protein	Active-Site Sequence	$E^{o\prime}$ of CXXC (V)	pK_a of CXXC	Primary role in the cell
Rat PDI	WCGHCK	–0.180[a]	6.7[b]	oxidase/isomerase
E. coli Trx	WCGPCK	–0.270[c]	7.5[d]	reductase
E. coli DsbA	FCPHCY	–0.120[e]	3.3[f]	oxidase

[a] Determined from the equilibrium constant with GSH/GSSG and Trx (Lundström & Holmgren, 1993).

[b] Determined from the rate of inactivation by alkylation (Hawkins & Freedman, 1991).

[c] Determined from the equilibrium constant for the thioredoxin reductase-catalyzed reaction with NADPH/NADP$^+$ (Krause *et al.*, 1991).

[d] Determined by ^{13}C-NMR spectroscopy for the state in which Asp26 and Cys35 are protonated (Chivers *et al.*, 1997b).

[e] Determined from the equilibrium constant with GSH/GSSG (Hennecke *et al.*, 1997; Huber-Wunderlich & Glockshuber, 1998).

[f] Determined by ultraviolet spectroscopy (Huber-Wunderlich & Glockshuber, 1998).

Table 1.2 Properties of active-site variants of thioredoxin

Protein	pK_a of CXXC	$E^{\circ\prime}$ of CXXC (V)	Relative doubling time of complemented *pdi1*Δ yeast
Yeast PDI	ND	ND	1.0
Rat PDI	6.7[a]	–0.180[b]	1.8 ± 0.2[c]
CGPC Trx	7.5[d]	–0.270[e]	NC
CGPS Trx	ND	–	4.3 ± 0.5[f]
CGHC Trx	ND	–0.235[e]	4.4 ± 0.8[f]
CVWC Trx	6.2[d]	–0.230[g]	3.8 ± 0.4[f]
CWGC Trx	6.1[d]	–0.200[g]	2.2 ± 0.2[f]

ND, not determined; NC, no complementation.

[a] Determined from the rate of inactivation by alkylation (Hawkins & Freedman, 1991).

[b] Determined from the equilibrium constant with GSH/GSSG and Trx (Lundström & Holmgren, 1993).

[c] Data from (Laboissière *et al.*, 1995b).

[d] Determined by ^{13}C-NMR spectroscopy for the state in which Asp26 and Cys35 are protonated (Chivers *et al.*, 1997a).

[e] Determined from the equilibrium constant of the thioredoxin reductase-catalyzed reaction with NADPH/NADP$^+$ (Krause *et al.*, 1991).

[f] Data from (Chivers *et al.*, 1996).

[g] Determined from the equilibrium constant of the thioredoxin reductase-catalyzed reaction with NADPH/NADP$^+$ (Chivers *et al.*, 1996).

Table 1.3 Properties of active-site variants of protein disulfide isomerase

PDI[a]	Dithiol oxidation activity[b]	Disulfide reduction activity[c]	Disulfide isomerization activity[d]	Doubling time of complemented *pdi1*Δ *S. cerevisiae*[e]
CGHC	100	100	100	1.8 ± 0.2
CGHS	3	6	93	2.3 ± 0.6
SGHC	0	3	4	NC

Data from (Laboissière *et al.*, 1995b).

NC, no complementation.

[a] For each protein, the sequence indicated is present in both active sites of rat PDI.

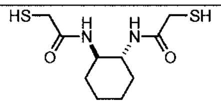
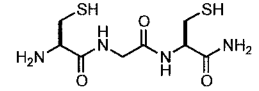
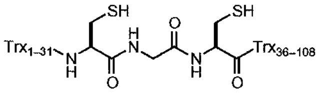
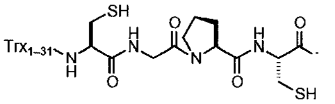
[b] Percentage of wild-type PDI activity for the activation of reduced RNase A.

[c] Percentage of wild-type PDI activity for the reduction of the disulfide bonds in insulin.

[d] Percentage of wild-type PDI activity for the activation of scrambled RNase A.

[e] Relative to cells complemented with *S. cerevisiae* PDI.

Table 1.4 Properties of small-molecule dithiol catalysts of disulfide isomerization

Molecule	Structure	First pK_a	$E^{o\prime}$ (V)	Specific Activity (U/mol)[a]
BMC		8.3[b]	−0.240[b]	56[b]
CGC		8.7[d]	−0.167[d]	132[c]
CGC Trx		ND	≥-0.200[e]	3300[c]
CGPC Trx		7.5[f]	−0.270[g]	0[c]

ND, not determined.

[a] One unit (U) of reduced catalyst activates 1 mmol of sRNase A per min in 0.1 M Tris–HCl buffer, pH 7.6, containing EDTA (1 mM).

[b] pK_a value was determined by ultraviolet spectroscopy; $E^{o\prime}$ value was determined by equilibration with β-mercaptoethanol/β-hydroxyethyldisulfide (Woycechowsky *et al.*, 1999).

[c] Data from (Woycechowsky & Raines, 2003).

[d] pK_a value was determined by ultraviolet spectroscopy; $E^{o\prime}$ value was determined by equilibration with β-mercaptoethanol/β-hydroxyethyldisulfide (Woycechowsky & Raines, 2003).

[e] Determined from the equilibrium constant for the thioredoxin reductase-catalyzed reaction with NADPH/NADP$^+$ (Woycechowsky & Raines, 2003).

[f] Determined by ^{13}C-NMR spectroscopy for the state in which Asp26 and Cys35 are protonated (Chivers *et al.*, 1997b).

[g] Determined from the equilibrium constant of the thioredoxin reductase-catalyzed reaction with NADPH/NADP$^+$ (Krause *et al.*, 1991).

Table 1.5 The effect of peptide chain length on K_m for binding to type I and type II prolyl 4-hydroxylase

Substrate	K_m (μM)	
	α(I)	α(II)
Ac-(Pro-Pro-Gly)	23,000[a]	n.d.
$(\text{Pro-Pro-Gly})_5$	180[b]	490[d]
$(\text{Pro-Pro-Gly})_{10}$	18[c]	95[c]
Protocollagen	0.2[c]	1.1[c]

n.d. not determined

[a] Data from (Prockop *et al.*, 1976).

[b] Data from (Myllyla *et al.*, 1977).

[c] Data from (Annunen *et al.*, 1997).

[d] Data from (Hieta *et al.*, 2003).

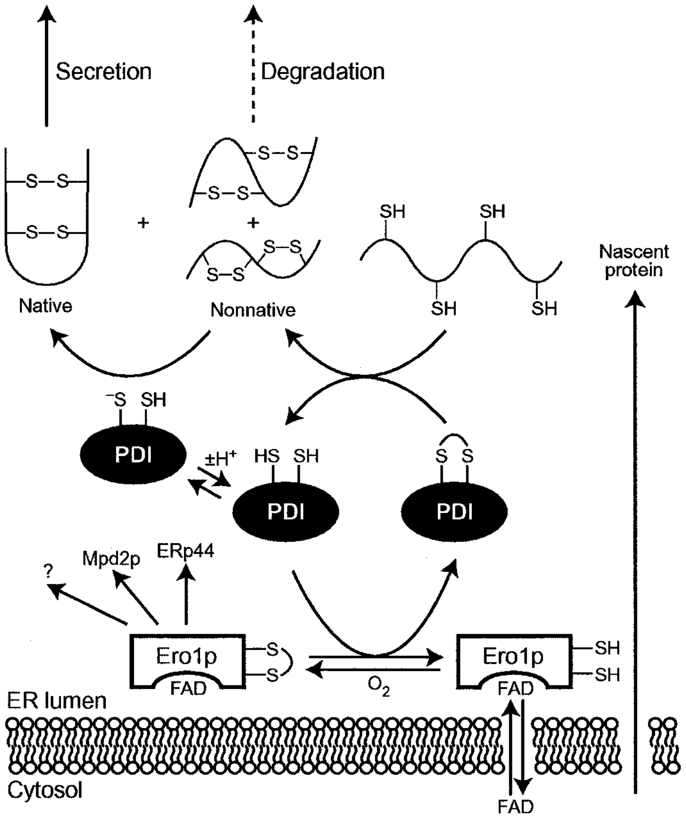

Secretion
Degradation
Native
Nonnative
Nascent protein
SH
-S SH
PDI
±H+
HS SH
S S
Mpd2p
ERp44
?
Ero1p
FAD
O_2
ER lumen
Cytosol

Figure 1.1 The pathway of native disulfide formation in the lumen of the ER. FAD-bound Ero1p (Tu *et al.*, 2000) (and presumably Ero1-Lα and Ero-Lβ in humans) specifically oxidizes PDI as well as ERp44 (human cells), Mpd2p (yeast), and perhaps other proteins (Frand & Kaiser, 1999; Anelli *et al.*, 2002). Ero1p uses molecular oxygen to reoxidize itself for further folding cycles (Tu & Weissman, 2002). Oxidized PDI catalyzes the formation of disulfide bonds in newly synthesized proteins; the thiolate form of reduced PDI catalyzes the isomerization of nonnative disulfide bonds (Laboissière *et al.*, 1995b). Proteins that do not achieve the native state are degraded rather than secreted. For simplicity, only one of the two active sites of Ero1p and PDI is shown.

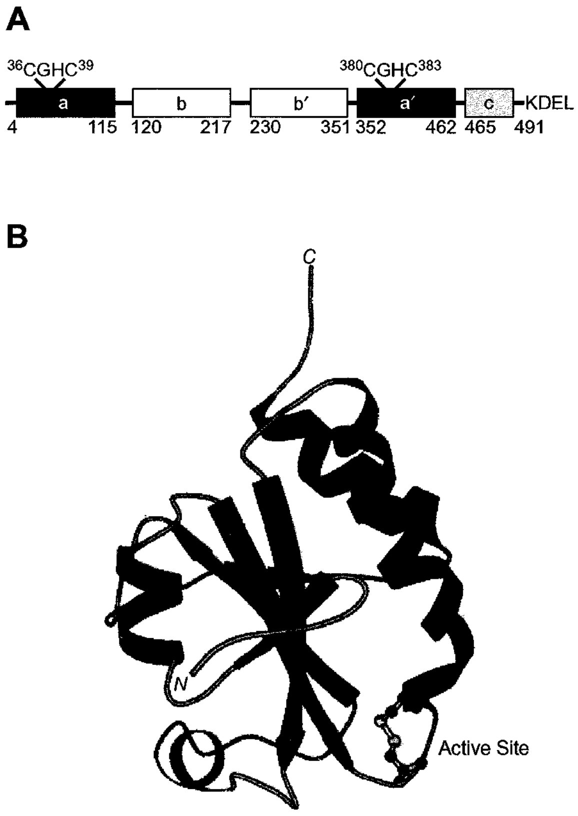

A
36CGHC39
380CGHC383
a
b
b′
a′
c
KDEL
4
115
120
217
230
351
352
462
465
491
B
C
N
Active Site

Figure 1.2 PDI is composed of four thioredoxin-like domains. **A**. The a and a′ domains each contain a CGHC active site. The b and b′ domains are likely involved in substrate binding. The c domain is cationic and ends with a *C*-terminal ER-retention sequence. The numbering system used here is based on the sequence for rat PDI (Darby *et al.*, 1999). **B**. The thioredoxin fold of the a domain of PDI is revealed in the NMR structure (Kemmink *et al.*, 1996). The CGHC active site is shown at the *N*-terminus of an α helix.

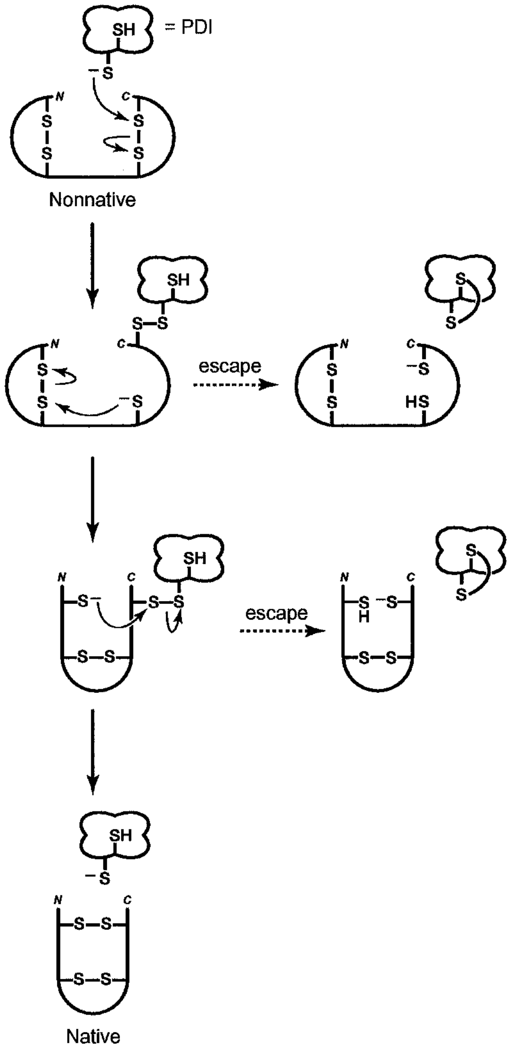


Figure 1.3 Mechanism of disulfide isomerization (Woycechowsky *et al.*, 1999). Isomerization begins with the nucleophilic attack of a thiolate provided by the catalyst (such as PDI) to form a mixed disulfide intermediate between catalyst and substrate. Further intramolecular thiol–disulfide interchange reactions are performed by substrate thiolates. If these interchange reactions are slow, then a second thiol in the catalyst can provide an escape route, releasing trapped intermediates by reduction of the mixed disulfide bond (Walker & Gilbert, 1997). Native disulfide bonds can then be formed by reoxidation.

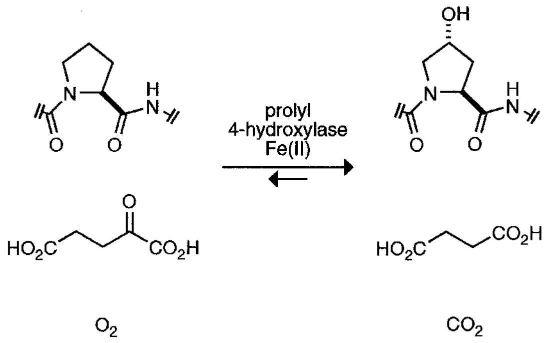

prolyl
4-hydroxylase
Fe(II)
OH
H
N
O
HO2C
CO2H
O2
CO2

Figure 1.4 Reaction catalyzed by prolyl 4-hydroxylase.

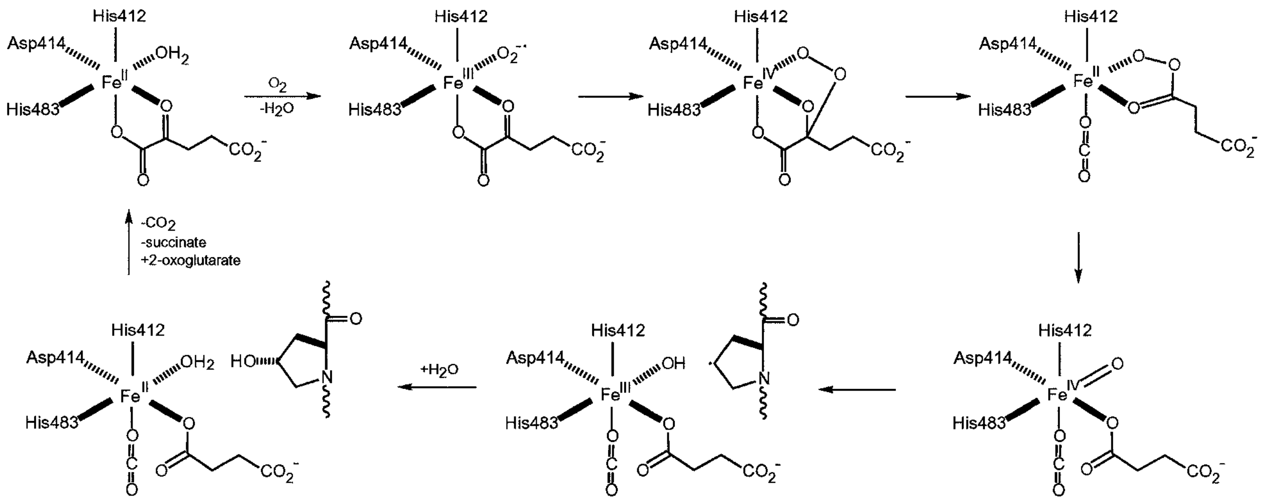

His412
Asp414
His483
Fe^{II}
Fe^{III}
Fe^{IV}
OH_2
$O_2^{\cdot-}$
OH
CO_2^-
O_2
$-H_2O$
$+H_2O$
$-CO_2$
-succinate
+2-oxoglutarate
HO
N

Figure 1.5 Putative mechanism for prolyl hydroxylation by P4H.

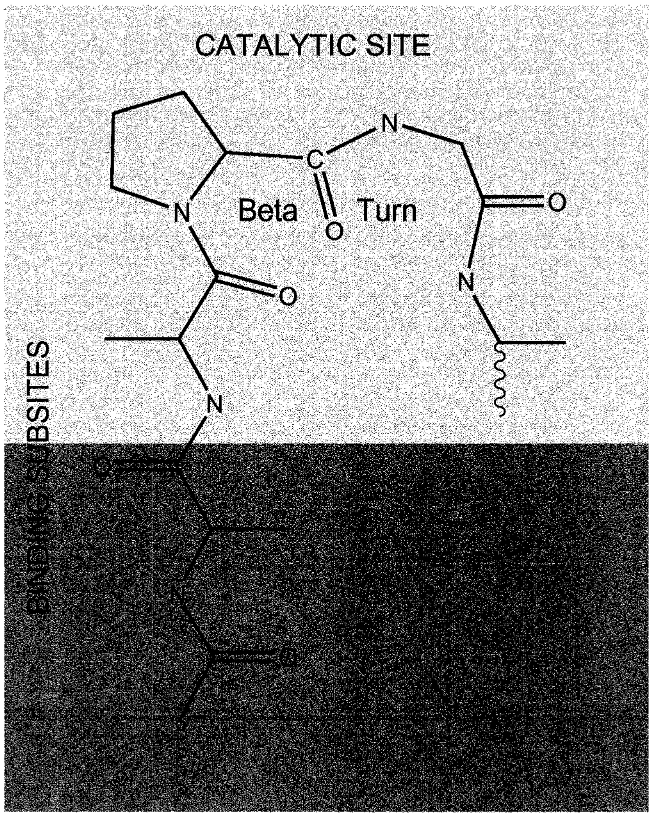

CATALYTIC SITE
Beta
Turn
BINDING SUBSITES
N
C
O

Figure 1.6 Proposed mode of peptide substrate binding to P4H (adapted from (Atreya & Ananthanarayanan, 1991)). A β-turn binds in the catalytic site for hydroxylation while the extended peptide chain (a polyproline-II helix) interacts with the enzyme at binding subsites.

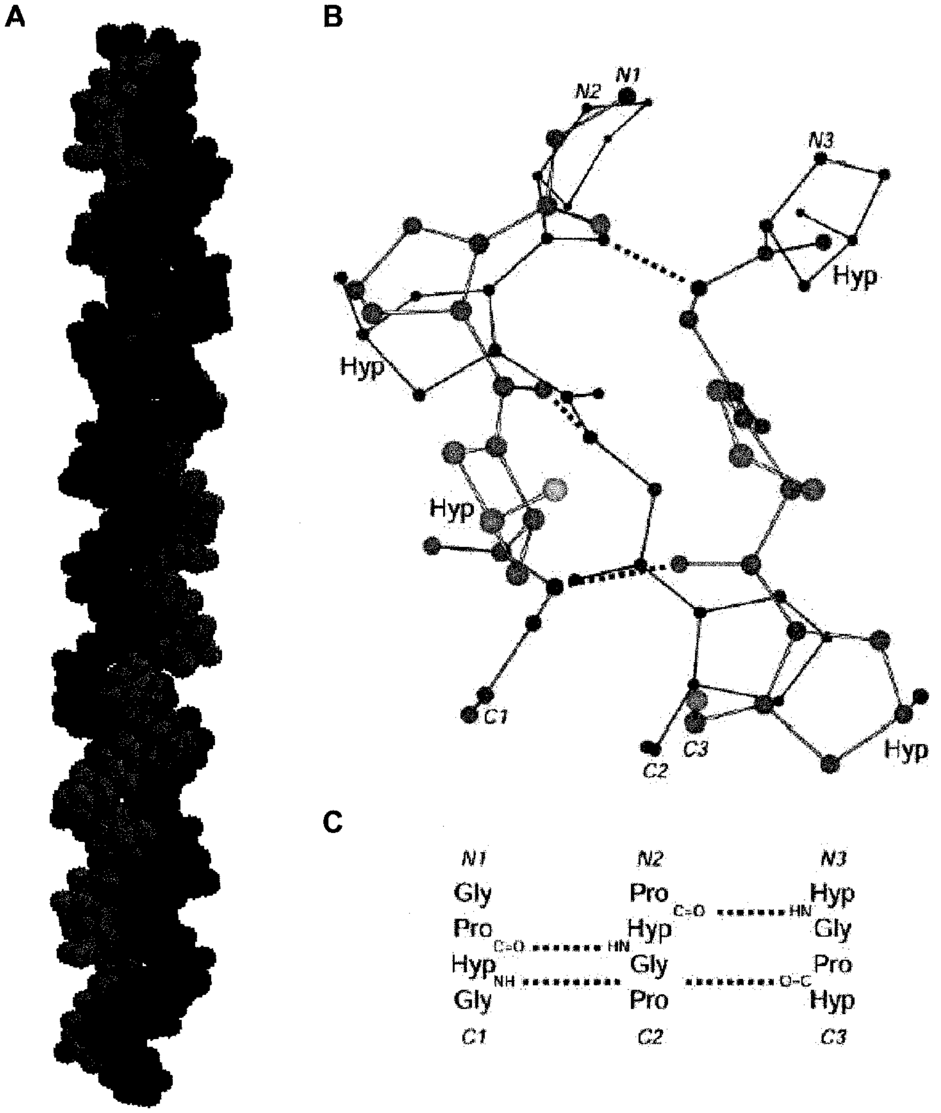

A
B
N1
N2
N3
Hyp
Hyp
Hyp
C1
C2
C3
Hyp
C
N1 N2 N3
Gly Pro Hyp
C=O HN
Pro Hyp Gly
C=O HN
Hyp Gly Pro
NH O-C
Gly Pro Hyp
C1 C2 C3

Figure 1.7 Structure of the triple helix of a collagen peptide mimic. **A.** Crystallographic structure of [(Pro–Hyp–Gly)$_4$–Pro–Hyp–Ala–(Pro–Hyp–Gly)$_5$]$_3$ with each strand indicated with a different color. Atomic coordinates are from (Bella *et al.*, 1994) (PDB entry 1CAG). **B.** Ball-and-stick representation of a segment of the structure in panel A, indicating (2*S*,4*R*)-hydroxyproline residues and backbone hydrogen bonds. **C.** Register of the residues in the three strands of panel B.

Chapter Two

Catalysis of Protein Disulfide Bond Isomerization

2.1 Abstract

Protein disulfide isomerase (PDI) catalyzes the rearrangement of nonnative disulfide bonds in the endoplasmic reticulum of eukaryotic cells, a process which is often rate-limiting for reaching the native protein conformation. Our understanding of the catalytic mechanism of disulfide bond isomerization has been limited by the available *in vitro* assays. Complex protein substrates do not permit direct observation of isomerization and are time-consuming to perform. The heterogeneity of these substrates also makes analysis of mechanistic details difficult. In the present study, we have developed a continuous assay for isomerization in which the substrate is a synthetic 18-residue peptide with two nonnative disulfide bonds. A variant of PDI in which the *C*-terminal active-site cysteine residues are replaced with alanine residues retains 50% of the activity of wild-type PDI in catalysis of disulfide bond isomerization. This result suggests that the *C*-terminal cysteines have a limited role in the isomerization of peptide disulfide bonds. Future mechanisic studies as well as characterization of small-molecule isomerases will benefit from this direct and rapid assay for disulfide bond isomerization.

2.2 Introduction

The native conformation of disulfide bonds in a protein is achieved by the oxidation of cysteine thiols followed by the rearrangement of nonnative disulfide bonds. In the endoplasmic reticulum (ER) of eukaryotic cells, native disulfide bond formation is catalyzed by protein disulfide isomerase (PDI; (Freedman *et al.*, 1994; Kersteen & Raines, 2003)). This 55-kDa enzyme contains four distinct structural domains (a, a′, b, and b′) with homology to thioredoxin and an acidic *C*-terminal domain (c) (Kemmink *et al.*, 1997; Darby *et al.*, 1999). The a and a′ domains each contain Cys-Gly-His-Cys (CGHC) active-site motifs, which catalyze thiol–disulfide interchange reactions leading to the oxidation of cysteine residues and the reduction and isomerization of disulfide bonds (Freedman *et al.*, 1994). The role of the b domain is unclear but likely includes maintaining the structural integrity of the enzyme (Koivunen *et al.*, 2005). The b′ domain does not contain catalytic residues but is required for substrate recognition and binding (Klappa *et al.*, 1998; Cheung & Churchich, 1999; Koivunen *et al.*, 2005).

PDI is an essential enzyme in *Saccharomyces cerevisiae* (Farquhar *et al.*, 1991; LaMantia *et al.*, 1991; Scherens *et al.*, 1991; Tachikawa *et al.*, 1991). Still, there has been much debate over which of its cellular roles is necessary for survival. A variant of rat PDI in which both active sites contain a CGHS sequence, instead of CGHC, restores viability to *pdi1*Δ *S. cerevisiae* (Laboissière *et al.*, 1995b). *In vitro*, this PDI variant is proficient in catalysis of disulfide isomerization but has low dithiol oxidation activity and low disulfide reduction activity (Laboissière *et al.*, 1995b; Walker *et al.*, 1996). Further, PDI homologs, including Eug1p and variants of thioredoxin, which are inefficient in catalysis of disulfide bond formation and reduction but can catalyze disulfide bond rearrangement, are able to rescue *pdi1*Δ *S. cerevisiae* (Tachibana & Stevens, 1992; Chivers *et al.*, 1996). Recently, however, Gilbert and coworkers have examined the effects of variable expression levels of yeast PDI variants on the growth and viability of *pdi1*Δ *S. cerevisiae* (Xiao *et al.*, 2001; Solovyov *et al.*, 2004; Xiao *et al.*, 2004). These studies suggest that it is the oxidase activity of PDI that limits growth and that isomerization might not be rate-limiting to the folding of proteins essential for yeast viability. Still, for many secretory proteins the isomerization of nonnative disulfide bonds is the rate-determining step in oxidative protein folding (Narayan *et al.*, 2000).

The mechanism by which PDI catalyzes disulfide bond isomerization involves the nucleophilic attack of the reactive *N*-terminal thiolate of the CGHC motif on a nonnative substrate disulfide to form a mixed disulfide intermediate (Figure 2.1). The resulting substrate thiolate is then free to attack a nonnative substrate disulfide. Further substrate rearrangements continue until the native disulfide pairings are reached and PDI is released without any net reduction or oxidation. Under some conditions, substrate thiolates could be incapable of completing the required disulfide rearrangements, in effect trapping an enzyme–substrate mixed disulfide. The *C*-terminal cysteine residue of the CGHC motif plays a critical role in rescuing these unproductive intermediates by forming an active-site disulfide bond and releasing a reduced substrate (Walker *et al.*, 1996; Walker & Gilbert, 1997). Cycles of reduction and reoxidation would ultimately produce a native protein structure.

Although PDI has been studied for decades, the catalytic mechanism of disulfide isomerization is not well understood. Questions remain regarding the recognition of misfolded substrates and the individual chemical steps that follow initial thiolate attack. The most common *in vitro* isomerization assays use heterogeneous substrates that make it difficult to decipher mechanistic details (Gilbert, 1998). For example, ribonuclease A (RNase A) is a 124-residue protein with four native disulfide bonds that are required for catalytic activity. Random oxidation of the eight cysteine residues under denaturing conditions gives scrambled RNase A (sRNase A), a mixture of up to 105 distinct fully oxidized species. Bovine pancreatic trypsin inhibitor (BPTI) contains only six cysteine residues but can still form 15 fully oxidized species. The pathway of disulfide bond isomerization in a redox buffer is complicated further by an even greater number of potential mixed-disulfide reaction intermediates.

In recent years, several *in vitro* assays for disulfide bond formation and reduction have been developed which use simple peptide substrates. Peptides such as somatostatin (a 14-residue, cysteine-containing peptide) and mastoparan (a 14-residue, cyteine-free peptide) bind reversibly to the same site on PDI as larger protein substrates including sRNase A (Klappa *et al.*, 1997). Substrate binding occurs predominantly in a small hydrophobic binding pocket in the b′ domain which is sufficient for the binding of small peptides like somatostatin (Pirneskoski *et al.*, 2004). Larger substrates like BPTI and sRNase A also require the a′ domain for efficient binding and catalysis (Klappa *et al.*, 1998).

Assays which use small peptides are useful because they simplify the catalytic cycle and make it possible to dissect individual mechanistic steps (Freedman *et al.*, 2003). The PDI-catalyzed oxidation of two cysteine residues in a 28-residue peptide has been studied using a discontinuous HPLC-based assay (Darby *et al.*, 1994). The substrate appears to have no secondary or tertiary structure and therefore is useful for the study of the chemical oxidation reaction without any conformational effects to complicate analysis. The technique of fluorescence resonance energy transfer (FRET) has been used in the development of continuous assays which monitor the formation and reduction of disulfide bonds. A 10-residue peptide using tryptophan as the fluorophore and arginine as the quencher shows a 19% decrease in fluorescence emission upon oxidation of the two cysteine residues (Ruddock *et al.*, 1996). This substrate has been used to analyze the catalytic properties of PDI and its homologs (Ruddock *et al.*, 1996; Alanen *et al.*, 2003; Lappi *et al.*, 2004; Pedone *et al.*, 2004). Another assay uses a substrate in which two peptides, one with an aminobenzoic acid fluorophore and the other with a nitrotyrosine quencher, are connected by a poly(ethyleneglycol) linker (Christiansen *et al.*, 2004).

A similar substrate for the *in vitro* study of disulfide bond isomerization is desirable. A peptide substrate with four cysteine residues would have only three possible fully oxidized species. One of these oxidized conformations should be the most thermodynamically stable and therefore "native", and it should be distinguishable from the "nonnative" conformations. To be a substrate for isomerization, the disulfide bonds in the native conformation should have a reduction potential (E^{o}) that is more negative (*i.e.* more stable) than the reduction potential of the PDI active sites ($E^{o} = -0.180$ V; (Lundström & Holmgren, 1993)). Then, PDI would not simply catalyze the reduction of the native disulfide bonds. Ideally, structural changes accompanying isomerization could be followed in a continuous manner.

We used tachyplesin I (TI) as a starting point for the development of such a substrate. TI is a 17-residue antimicrobial peptide (Figure 2.2) that has been isolated from the hemocytes of the horseshoe crab, *Tachypleus tridentatus* (Nakamura *et al.*, 1988). NMR structural analysis has revealed that the peptide folds into an antiparallel β-sheet composed of residues 3–8 and 11–16 Residues Arg9 and Gly10 define a β turn, and the *N*- and *C*-terminal residues are disordered (Figure 2.2; (Kawano *et al.*, 1990; Laederach *et al.*, 2002)). The β-sheet is stabilized by two disulfide bonds and six hydrogen bonds. The ability of TI to permeabilize membranes through the formation of anion-selective pores is attributed to an amphiphilic structure composed of a large number of hydrophobic and cationic residues (Matsuzaki *et al.*, 1997; Oishi *et al.*, 1997; Laederach *et al.*, 2002).

By incorporating a dansyl fluorophore near the *C*-terminus of the TI peptide and selectively forming nonnative disulfide bonds, we have developed a substrate for a continuous FRET assay that monitors the isomerization of nonnative disulfide bonds. We have used this assay to compare active-site variants of PDI. The results provide new insight into the mechanism of disulfide bond isomerization.

2.3 Materials and Methods

Materials. Escherichia coli strains BL21(DE3) and DH5α, and the pET22b(+) expression vector were from Novagen (Madison, WI). DNA oligonucleotides for mutagenesis and sequencing were from Integrated DNA Technologies (Coralville, IA). Enzymes for DNA manipulation were from Promega (Madison, WI) and New England Biolabs (Beverly, MA). DNA sequencing reactions were performed with the BigDye kit from Applied Biosystems (Foster City, IA) and CleanSeq magnetic beads for reaction clean-up were from Agencourt Bioscience (Beverly, MA). DNA sequences were determined on an Applied Biosystems automated sequencing instrument at the University of Wisconsin Biotechnology Center. Poly(cytidylic acid) (poly(C)) was from Midland Certified Reagents (Midland, TX) and was precipitated from aqueous ethanol (70% v/v) before use. Fmoc-protected amino acids were from Novabiochem (La Jolla, CA). Fmoc-Lys(dansyl)-OH was from AnaSpec (San Jose, CA). All other reagents were of reagent grade or better and used without further purification.

Luria–Bertani (LB) medium contained (in 1.0 L) tryptone (10 g), yeast extract (5 g), and NaCl (10 g). Terrific broth (TB) medium contained (in 1.0 L) tryptone (12 g), yeast extract (24 g), glycerol (4 mL), K_2HPO_4 (72 mM), and KH_2PO_4 (17 mM). All media were prepared in deionized, distilled water, and autoclaved.

Instrumentation. Measurements of UV and visible absorbance were made with either a Cary model 50 or a Cary model 3 spectrophotometer equipped with a Cary temperature controller (Varian, Palo Alto, CA). Peptide synthesis was conducted with a Pioneer automated synthesizer (PerCeptive Biosystems) 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 717plus autosampler, and a 996 photodiode array detector. Fast performance liquid chromatography (FPLC) was performed with an ÄKTA system from Amersham Pharmacia (Piscataway, NJ). Fluorescence was measured with a QuantaMaster 1 photon-counting fluorescence spectrometer equipped with stirring (Photon Technology International, South Brunswick, NJ). Matrix-assisted laser desorption/ionization time of flight (MALDI–TOF) mass spectrometry was performed on a PerkinElmer (Wellesley, MA) Voyager MALDI–TOF mass spectrometer at the University of Wisconsin Biophysics Instrumentation Facility.

Plasmid for Production of Human PDI and Variants. A plasmid containing the human PDI gene was a generous gift from Alan D. Attie (University of Wisconsin–Madison). This plasmid contains the gene for a variant of human PDI with two SGHC active sites (PDIx) inserted between the *Bgl*II and *Bam*HI sites of the pET22b(+) vector. The PDIx gene was isolated by restriction digest and inserted into the *Bam*HI site of pBK1 (a pET22b(+) vector in which the

EcoRI site had been removed) to give pBK1.hPDIx1. Two unique restriction sites — *Eco*RI, 5′ to the first active site, and *Mfe*I, 3′ to the second active site — were added to the PDI gene using the Kunkel method of mutagenesis to give pBK1.hPDIx4 (Kunkel, 1985). A unique *Nde*I site was added at the 5′ end of the coding region of the gene, placing an ATG start codon in place of the final codon of the signal sequence coding region. Restriction digest with *Nde*I and *Bam*HI, followed by ligation, removed the 5′ signal-sequence coding region to give pBK1.hPDIx6, which contains human PDI coding for two SGHC active sites inserted between the *Nde*I and *Bam*HI restriction sites of pBK1.

A vector for the production of wild-type human PDI was obtained by PCR mutagenesis with two primers—one complementary to the 5′ active site region, including the *Eco*RI restriction site (GTGGAATTCTATGCTCCTTGGTGCGGCCACTG), and one complementary to the 3′ active site region, including the *Mfe*I restriction site (GGGAGCCAATTGTTTGCAGTGACCACACCATGG). The PCR product was cloned into the PCR4-TOPO cloning vector (Invitrogen, Carlsbad, CA) and confirmed by DNA sequencing. Digestion with *Eco*RI and *Mfe*I followed by ligation gave pBK1.hPDI1, the wild-type human PDI gene in pBK1.

A vector for the production of the CGHA:CGHA PDI variant (in which both active sites have a CGHA sequence) was obtained in a series of steps. First, QuikChange (Stratagene, La Jolla, CA) site-directed mutagenesis was used to change the second cysteine codon of the *N*-terminal active site to an alanine codon (GCT). The Kunkel method of mutagenesis was then used to change the *C*-terminal cysteine codon of the second active site to an alanine codon (GCT). The resulting plasmid, pBK1.PDI2, codes for human PDI with two CGHA active sites.

Expression and Purification of PDI. The pBK1.hPDI1 (or plasmid for an appropriate PDI variant) was transformed by electroporation into the *E. coli* strain BL21(DE3) and grown on LB agar plates containing ampicillin (200 μg/mL). A starter culture in LB medium (25 mL), supplemented with ampicillin (200 μg/mL), was inoculated with a single colony and grown overnight at 37 °C. Overnight cultures were used to inoculate TB medium (1.0 L) containing ampicillin (200 μg/mL) to *OD* = 0.01 at 600 nm. The inoculated culture was shaken (210 rpm) at 37 °C until reaching *OD* = 1.7–2.0 at 600 nm. Protein expression was induced by the addition of isopropyl-1-thio-β-D-galactopyranoside (IPTG, 0.5 mM) and shaking for 4 h at 37 °C.

Cells (10 g wet weight) were harvested by centrifugation and resuspended in 15 mL of 50 mM Tris–HCl buffer, pH 8.0 containing EDTA (2 mM). Cells were lysed by passage three times through a French pressure cell. Insoluble material was removed by ultracentrifugation at 50,000*g* for 1 h at 4 °C. The lysate was clarified further by an ammonium sulfate fractionation. Saturated ammonium sulfate was added to the lysate to 55% saturation, and the pellet was removed by centrifugation at 15,000*g* for 30 min. The supernatant was precipitated by adding saturated ammonium sulfate to 85% saturation. The pellet was isolated by centrifugation at 15,000*g* for 30 min and resuspended in 25 mM sodium phosphate buffer, pH 8.0, containing NaCl (0.40 M). The 55–85% ammonium sulfate fraction was dialyzed overnight against 4.0 L of the same buffer.

The dialyzed lysate was loaded onto a Hi Load 26/60 Superdex-200 gel filtration column (Amersham Biosciences) that had been preequilibrated with 25 mM sodium phosphate buffer, pH 8.0, containing NaCl (0.40 M). The column was eluted with the same buffer at a flow rate of 2.0 mL/min. Fractions containing PDI were identified by SDS–PAGE analysis and combined. The elution of PDI from the gel filtration column was indicative of a protein of 58 kDa, based on a calibration curve obtained with a calibration kit (29–700 kDa) from Sigma Chemical (St. Louis, MO). Fractions containing PDI were combined and dialyzed overnight against 4.0 L of 25 mM sodium phosphate buffer, pH 6.8.

Dialyzed protein from gel filtration was injected at 2 mL/min onto a column (8 mL) of Resource Q anion-exchange resin (Amersham Biosciences) that had been preequilibrated with 25 mM sodium phosphate buffer, pH 6.8. After washing with 24 mL of equilibration buffer, PDI was eluted with a linear gradient (90 + 90 mL) of NaCl (0–1.00 M) in equilibration buffer. Fractions containing PDI, which eluted at approximately 0.32 M NaCl, were combined and dialyzed overnight against 4.0 L of 25 mM sodium phosphate buffer, pH 6.0. After dialysis, purified protein was concentrated to approximately 1.0 mg/mL using a Vivaspin concentrator (20 mL, molecular mass cutoff: 10 kDa, Vivascience AG, Hanover, Germany) and stored at –80°C. Typically, 10 mg of PDI (>95% pure) was obtained from a 1-L growth.

Synthesis of Fluorescent Tachyplesin. The sequence of wild-type tachyplesin I was altered by changing Lys1 to Gly, Arg17 to Lys, and adding a Gly residue at the *C*-terminus (Figure 2.2). The synthesis of native peptide (nTI) was performed using standard Fmoc-protection strategies. The *N*-terminus of the peptide was capped with an acetyl group. All cysteine residues were protected with trityl (Trt) groups.

The nTI peptide (approximately 0.1 mmol) was cleaved from the solid support and its protecting groups were removed at room temperature for 1.5 h with a cocktail (5 mL) of trifluoroacetic acid (TFA, 82.5% v/v), phenol (5% v/v), thioanisole (5% v/v), water (5% v/v), and 2-ethanedithiol (EDT, 2.5% v/v). The peptide was precipitated by dripping into cold ethyl ether (30 mL) and collected by centrifugation. The peptide was washed two times with ethyl ether (30 mL) and the crude product was dried under vacuum.

The crude peptide was purified by reversed-phase HPLC using a Dynamax Microsorb preparative C18 column (21.4 mm x 250 mm), equipped with a guard column, from Varian (Lake Forest, CA). Reduced peptide was eluted with a linear gradient (40 min) of aqueous acetonitrile (25–65% v/v) containing TFA (0.1% v/v) at a flow rate of 14.6 mL/min. The native disulfide bonds were formed by incubation with aqueous dimethylsulfoxide (DMSO, 20% v/v) overnight at room temperature. Oxidized nTI was purified by HPLC as described above for reduced nTI (MALDI–MS $[M+H]^+$ calcd 2265.0; found 2265.8).

A dansyl (dns) fluorophore was attached to the ε-amino group of Lys17. Oxidized nTI (7.0 mg, 1 mM) in DMF (3.0 mL total) was stirred with DIPEA (100 mM, 100 equiv) at 4 °C for 5 min and was followed by the addition of dansyl chloride (1 mM, 1 equiv). The reaction was allowed to proceed for 6 h at 4 °C. The reaction was quenched by the addition of water (27 mL). Dansylated nTI (dns-nTI) was purified by HPLC on the preparative C18 column as indicated above for reduced nTI. The purified peptide (5.0 mg, 2%) was dried under vacuum and analyzed by MALDI mass spectrometry ($[M+H]^+$ calcd 2497.1; found 2497.0).

Scrambled tachyplesin (sTI) contains the same amino acid sequence as does nTI, but the pairing of disulfide bonds is nonnative. In sTI, Cys3 and Cys7 form one disulfide and Cys12 and Cys16 form a second disulfide. The synthesis of sTI was performed using standard Fmoc-protection strategies. The *N*-terminus of the peptide was capped with an acetyl group. The nonnative disulfide bonds were formed through a pairwise protecting group strategy (Figure 2.3). Cys3 and Cys7 were protected with acetamidomethyl (Acm) groups while Cys12 and Cys16 were protected with trityl (Trt) groups. Fmoc-Lys(dansyl) was coupled as residue 17.

Upon cleavage of the dns-sTI peptide (approximately 0.1 mmol) from the solid support, all protecting groups except Cys(Acm) were removed at room temperature for 1.5 h with a cleavage cocktail (10 mL) of TFA (91.5% v/v), phenol (2.5% v/v), water (2.5% v/v), EDT (2.5% v/v), and triisopropylsilane (1.0% v/v). The peptide was precipitated by dripping into cold ethyl ether (30 mL) and collected by centrifugation. The crude peptide (dns-sTI(Acm)) was washed two times with ethyl ether (30 mL) and dried under vacuum.

HPLC on a Dynamax Microsorb preparative C18 (21.4 x 250 mm) column was used to purify crude dns-sTI(Acm). Peptide was eluted with a linear gradient (30 min) of acetonitrile (30–60% v/v in water) containing TFA (0.1% v/v). The first nonnative disulfide bond was formed by oxidation of dns-sTI(Acm) with aqueous DMSO (20% v/v) at room temperature overnight. Oxidized dns-sTI(Acm) was purified by preparative HPLC by the same method described above for reduced dns-sTI(Acm). Purified peptide was dried by lyophilization (MALDI–MS $[M+H]^+$ calcd 2641.2; found 2640.9).

The second nonnative disulfide bond (Cys3–Cys7) was formed by the method of Fujii and coworkers (Tamamura *et al.*, 1993; Tamamura *et al.*, 1995; Tamamura *et al.*, 1998). The Acm groups were removed by reaction of peptide (8 μmol) with silver trifluoromethanesulfonate (AgOTf, 50 equiv) in TFA (2 mL) containing anisole (1% v/v) at 4 °C for 1.5 h. Peptide was precipitated in cold ethyl ether (5 mL), and collected by centrifugation, and washed with cold ethyl ether (2 x 5 mL). Residual ether was removed under vacuum. The *N*-terminal disulfide bond was formed by oxidation with DMSO (50% v/v) in 1 N HCl (4 mL) at room temperature for 7 h. Silver chloride was removed by filtration and water was added (6 mL). The product peptide (dns-sTI) was purified by HPLC on a Zorbax StableBond reversed-phase C8 column (9.4 x 250 mm) from Agilent (Wilmington, DE). Peptide was eluted with a linear gradient (40 min) of acetonitrile (27–37% v/v in water) containing TFA (0.1% v/v) at a flow rate of 4.1 mL/min. Dns-nTI elutes at 31.1% acetonitrile while dns-sTI elutes at 31.7% acetonitrile. The nonnative peptide (1.4 mg, 0.6 %) was dried under vacuum, its purity confirmed by analytical HPLC (98%), and its identity confirmed by MALDI mass spectrometry ($[M+H]^+$ calcd 2497.1; found 2497.4).

Physical Characterization of Dns-TI. The extinction coefficient for dns-nTI was determined at 280 nm by the method of Gill and von Hippel (Gill & von Hippel, 1989). The extinction coefficient at 280 nm for the dansyl fluorophore was estimated (ε_{280} = 4141 $M^{-1}cm^{-1}$) using known concentrations of dansyl glycine (Sigma Chemical). The value obtained for dns-nTI (ε_{280} = 10680 $M^{-1}cm^{-1}$) was assumed to be the same for dns-sTI, and was used to calculate peptide concentration in purified samples.

The disulfide bond arrangement in dns-nTI and dns-sTI were confirmed by proteolytic digestion with α-chymotrypsin. Peptide (15 μg) was incubated with α-chymotrypsin (0.5 μg) in Tris–HCl buffer (100 mM, pH 7.6) containing $CaCl_2$ (1 mM) at 37 °C for 18 h. Digested fragments were isolated by analytical HPLC on a reversed-phase C18 column (4.8 x 250 mm) from Varian, and were identified by MALDI mass spectrometry. Based on preliminary work with non-fluorescent nTI and sTI, we were aware of some unusual digestion patterns for the tachyplesin peptides (S. Barrows, personal communication). First, chymotrypsin cleaves both peptides after consecutive Arg residues at positions 14 and 15. Unusual interactions between chymotrypsin and consecutive Arg residues in a substrate peptide had been observed previously (Rabbani *et al.*, 1982). In addition, chymotrypsin is unable to cleave after Phe4 in the nonnative conformation. This position is located within the *N*-terminal disulfide bonded loop, which could hinder the accessibility of this amide bond to the protease. Based on these previous observations, the predicted fragmentation pattern for dns-nTI includes peptides with $[M+H]^+$ = 806.3 and 1148.5. Digestion of the purified peptide product from the dns-nTI synthesis yielded peptide products with $[M+H]^+$ = 806.5 and 1148.9, as observed by MALDI mass spectrometry. The predicted fragmentation pattern for dns-sTI includes peptides with $[M+H]^+$ = 788.3 and 1148.5. Digestion of the peptide product from the dns-sTI synthesis yielded peptide products with $[M+H]^+$ = 789.0 and 1149.3, as observed by MALDI mass spectrometry.

Fluorescence Detection. Fluorescence measurements were complicated by the adsorption of dns-nTI to typical cuvettes. This adsorption was lessened by the use of methacrylate cuvettes (4.5 mL, 1 x 1 cm, Fisher Scientific, Pittsburgh, PA). Cuvettes and stir bars were soaked in aqueous poly(ethylenimine) (PEI, 1% w/v) for at least 1 h. The cells were then rinsed with deionized, distilled water prior to performing fluorescence experiments. The reaction buffer (which was 100 mM Tris–HCl buffer, pH 7.6, containing 1 mM EDTA) was supplemented with IGEPAL CA-630 (0.01% w/v) to reduce peptide adsorption further. The reaction buffer was degassed under vacuum for 30 min, and flushed with argon for 40 min to minimize air oxidation. Assays were performed in a volume of 2.0 mL at room temperature. The excitation and emission slit widths were 4 nm each.

Fluorescence Isomerization Assays. Isomerization of dns-sTI to dns-nTI is accompanied by the formation of the native β-sheet conformation. For each reaction, dns-sTI substrate peptide (1.1 μM) was incubated in reaction buffer for 1 min to obtain an initial fluorescence value. PDI was preincubated with glutathione (116 μM GSH and 4 μM GSSG) in reaction buffer at room temperature for 30 min. To initiate the isomerization reaction, GSH (116 μM), GSSG (4 μM), and preequilibrated PDI (10–50 nM) were added simultaneously to the substrate. Isomerization was measured as a function of time by monitoring the increase in fluorescence emission at 465 nm upon excitation at 280 nm. A typical data set is shown in Figure 2.4B.

Kinetic parameters were determined with either nonlinear least-squares regression analysis using equation **2.1** for data from reactions carried to completion or linear least-squares regression analysis of initial velocity data using equation **2.2**.

$$F = F_0 - (F_{\max} - F_0)(1 - e^{-kt}) \qquad \mathbf{2.1}$$

$$F = F_0 + (F_{\max} - F_0)kt \qquad \mathbf{2.2}$$

The fluorescence emission (F) at a given time is measured as counts per second (c.p.s.). The substrate fluorescence (F_0) was measured before the addition of enzyme. The product fluorescence ($F_{\max}$) was determined by a fit of the data to equation **2.1**. The first-order rate constant k is equal to $(k_{cat}/K_m)[E]$ and was corrected for the uncatalyzed reaction. Values derived with equations **2.1** and **2.2** are within error. Rapid reaction of enzyme with substrate allows completion of the isomerase reaction. We therefore used equation **2.1** to derive the reported k_{cat}/K_m values for dns-sTI isomerization. Reactions were performed at substrate concentrations below the K_m. We were limited in the amount of substrate we could use in the assay by the sensitivity of the fluorescence spectrometer. Dns-sTI concentrations up to 16 μM did not saturate the enzyme (data not shown).

Determination of Disulfide Reduction Potential of nTI. The reduction potential of dns-nTI was determined by using fluorescence spectroscopy. Thiol–disulfide exchange equilibria were established between dns-nTI and glutathione. Reaction buffer containing IGEPAL CA-630

(0.01% v/v) was first degassed (30 min) and flushed with argon (> 40 min). Peptide (480 nM) was then added to the degassed buffer and various ratios of reduced to oxidized glutathione ($E^{o\prime}$ = –0.252 V; (Lees & Whitesides, 1993)). Fluorescence was measured once equilibrium had been reached (20 min). Fluorescence emission intensity was plotted as a function of the reduction potential of the buffer (Figure 2.5). The standard reduction potential of nTI was calculated with nonlinear least-squares analysis of the data according to the appropriate form of the Nernst equation:

$$F = F_0 + (F_{\max} - F_0) / (1 + e^{-(E-E^{\circ})(n\mathcal{F}/RT)}) \qquad \textbf{2.3}$$

where n is the number of electrons (2), $\mathcal{F}$ is the Faraday constant (96,494 J V^{-1} mol^{-1}), R is the gas constant (8.314 J K^{-1} mol^{-1}), T is the temperature (298.15 K), E is the reduction potential of the solution, and E° is the reduction potential of dns-nTI. To investigate cooperativity between the two disulfide bonds, the data were depicted in a Hill plot (Figure 2.5 inset), which here is a log plot of the ratio of reduced to oxidized peptide (determined by relative fluorescence emission intensity) as a function of $[GSH]^2/[GSSG]$. Linear least-squares analysis according to equation **2.4** was used to obtain the Hill coefficient of cooperativity (h) and to confirm the value of E° for dns-nTI that was calculated from the equilibrium constant (K_{eq}) according to the familiar form of the Nernst equation.

$$\log([\text{dns-rTI}]/[\text{dns-nTI}]) = h \log([GSH]^2/[GSSG]) - \log(K_{eq}) \qquad \textbf{2.4}$$

sRNase A Isomerization Assays. sRNase A was obtained by air oxidation of reduced RNase A (~35 μM) over several days in 100 mM Tris–HCl buffer, pH 8.0, containing guanidine–HCl (6.0 M) and EDTA (1.0 mM). The protein was then dialyzed against 100 mM Tris–HCl buffer, pH

7.6, containing EDTA (1.0 mM). This process produced RNase A that contained <0.1 mol thiol/mol protein. Isomerization of sRNase A was measured essentially as described previously (Laboissière *et al.*, 1995a; Woycechowsky *et al.*, 1999). To prevent air oxidation, reaction buffer (which was 100 mM Tris–HCl buffer, pH 7.6, containg EDTA (1 nM)) was degassed under vacuum for 30 min and flushed with argon for 30 min. Reactions (1.0 mL) were performed under a positive pressure of argon at 23 °C in reaction buffer containing GSH (1.0 mM), GSSG (0.2 mM) and PDI (360 nM). Addition of sRNase A (3.6 μM) was used to initiate the reaction. Aliquots (50 μL) were removed and quenched with acetic acid (5 μL, 10% v/v) at timed intervals. These aliquots were assayed for poly(C)-cleavage activity by monitoring the change in absorbance at 238 nm. Data were fitted to an exponential curve to obtain second-order rate constants. Specific activity was calculated from the difference in rate constants for the catalyzed and uncatalyzed reactions.

2.4 Results

Design and Synthesis of Substrate. A 17-residue peptide was used as the basis for a disulfide bond isomerization substrate. Its small size makes it simple to synthesize by solid-phase methods. Its two disulfide bonds lend stability to its native antiparallel β-sheet. We designed a FRET assay that makes use of Trp2 as the fluorescence donor and a dansyl moiety as a fluorescence acceptor. The Trp–dansyl interaction is sensitive enough to measure distance distributions within conformationally flexible proteins and peptides (Lakowicz *et al.*, 1990; Eis & Lakowicz, 1993; Maliwal *et al.*, 1993; Lakowicz *et al.*, 2004). It seemed likely that the relatively short Förster radius (approximately 20 Å (Lakowicz *et al.*, 1990)) for this FRET pair would allow us to distinguish between the native β-sheet conformation, in which the peptide termini are close together, and a nonnative conformation in which the termini are far apart.

Several changes were made to the native amino acid sequence to allow for the specific placement of fluorophores (Figure 2.2). All of the changes were at the *N*- and *C*-termini of the peptide, which are mobile in solution (Kawano *et al.*, 1990; Laederach *et al.*, 2002), and thus unlikely to contribute to β-sheet stability. Our design called for a dansyl fluorophore to be attached to a lysine residue at the *C*-terminus. In order to make a *C*-terminal lysine the only amine available for attachment of dansyl, we changed Lys1 to glycine and acetylated the *N*-terminus. We then installed a lysine residue at position 17 in place of the wild-type arginine residue. Finally, to facilitate peptide synthesis, we added a glycine residue following Lys17 in place of the native carboxamide.

The synthesis of nonnative disulfide bonds was achieved by a pairwise protecting group strategy in which the two *N*-terminal cysteine residues were Acm-protected during peptide synthesis while the two *C*-terminal cysteine residues were Trt-protected (Figure 2.3). The acid-labile Trt groups were removed during cleavage from the resin, and the *C*-terminal cysteines were oxidized by aqueous DMSO. Under these conditions the Acm groups remain intact as confirmed by MALDI mass spectrometry. In the final step of dns-sTI synthesis, the Acm groups were removed using AgOTf and the *N*-terminal cysteine residues were oxidized in DMSO/aqueous HCl (Tamamura *et al.*, 1993; Tamamura *et al.*, 1995; Tamamura *et al.*, 1998). Other oxidation methods, including silyl chloride/diphenylsulfoxide (Akaji *et al.*, 1992; Akaji, 1997) and TFA/DMSO (Cuthbertson & Indrevoll, 2000) as well as iodine oxidation (Kamber *et al.*, 1980; Sieber *et al.*, 1980), were unsuccessful due to modification and decomposition of Trp2. The nonnative conformation comprised 42% of the peptide product after oxidation by the

AgOTf/DMSO/aq. HCl method, with 58% being in the native conformation, as measured by integration of HPLC peaks. The overall yield for the synthesis of dns-sTI was less than 1% (1.4 mg from 100 μmol peptide synthesis) while the overall yield for the synthesis of dns-nTI, which did not require a pairwise cysteine protection, but did involve post-synthesis addition of the dansyl group, was 2% (5.0 mg from 100 μmol peptide synthesis).

Fluorescence Properties of Dns-nTI and Dns-sTI. In 100 mM Tris–HCl buffer, pH 7.6, containing EDTA (1 mM), dns-nTI (1.1 μM) has a maximal fluorescence emission upon excitation at 280 nm at 465 nm due to FRET between Trp2 and dansyl (Figure 2.4A). The same concentration of dns-sTI has a fluorescence emission spectrum that does not show significant fluorescence emission at 465 nm, but instead displays fluorescence emission from Trp2 at 340 nm (Figure 2.4A). The conversion from the nonnative to the native disulfide conformation is accompanied by an increase in fluorescence emission at 465 nm with $F_{\mathrm{max}}/F_0 = 28 \pm 2$ (after correcting for background fluorescence of the buffer, Figure 2.4B). Replicate preparations of peptides gave spectroscopic parameters that did not differ significantly.

Reduction Potential of Dns-nTI. The reduction potential (E^{o}) of a disulfide bond is a measure of its stability towards reduction. The dithiol of a more stable disulfide bond (more negative E^{o}) is a powerful reductant, and a less stable disulfide bond (more positive E^{o}) is a more powerful oxidant. The β-sheet conformation likely lends stability to the disulfide bonds of dns-nTI. We measured the reduction potential of dns-nTI by equilibration with varying ratios of reduced to oxidized glutathione which corresponded to different reduction potentials (Figure 2.5). The fluorescence emission intensity obtained in different redox buffers indicates the amount of oxidized versus reduced peptide. Nonlinear least-squares analysis of the data according to the Nernst equation gives $E^{o} = -0.313$ V (correlation coefficient: 0.990) for dns-nTI. A Hill plot of the data (Figure 2.5 inset) gives $h = 1.018$, suggesting that there is no significant cooperativity in the reduction of the two disulfide bonds. In addition, the data from this curve gives a value for E^{o} for dns-nTI (–0.314 V) which is in good agreement with that obtained above.

Isomerization of Dns-sTI by PDI. Disulfide bond isomerization in the presence of substoichiometric amounts of PDI was assayed by monitoring the increase in fluorescence emission at 465 nm upon excitation at 280 nm over time. In reaction buffer alone (no catalyst or redox buffer) the fluorescence emission of dns-sTI did not change over the time course of the assay (30 min, data not shown). The presence of glutathione (116 μM GSH and 4 μM GSSG) in the assay buffer ensured that PDI was maintained in an optimum redox state. Although air oxidation was minimized by degassing all buffers prior to assay, our fluorescence spectrometer system was not equipped for reaction under oxygen-free conditions. Attempts to remove the redox buffer from the reaction, using reduced PDI as catalyst, were unsuccessful due to the rapid oxidation of PDI before complete isomerization was achieved. In the absence of enzyme, glutathione alone efficiently converts dns-sTI to dns-nTI (Figure 2.6). At the concentrations used here, isomerization by glutathione shows a delay of approximately 20 s. PDI-catalyzed isomerization was corrected for nonenzymatic isomerization by subtracting the first-order rate constants obtained using equation **2.1**. The rate constant for the nonenzymatic reaction was calculated from the portion of the curve following the lag.

In the presence of wild-type PDI (20 nM), the conversion of dns-sTI to dns-nTI was nearly complete within approximately 7 min, as observed by fluorescence emission (Figure 2.6). In the presence of increasing concentrations of enzyme, the first-order rate constant increased as expected (data not shown). Using equation **2.1**, data from triplicate reactions and at varying concentrations of wild-type PDI give a k_{cat}/K_m value of $(1.7 \pm 0.5) \times 10^5\ M^{-1}\ s^{-1}$ (Table 2.1). This reaction is significantly more efficient than the oxidative folding of rRNase, which displays a k_{cat}/K_m value of $(1.2 \pm 0.5) \times 10^3\ M^{-1}s^{-1}$ (Walker *et al.*, 1996). The k_{cat}/K_m value obtained for the CGHA:CGHA variant of human PDI was $(8.6 \pm 0.35) \times 10^4\ M^{-1}\ s^{-1}$, approximately half that of the wild-type enzyme (Table 2.1).

Isomerization of sRNase A by PDI. The traditional assay for disulfide bond isomerization measures the increase in ribonuclease activity upon addition of catalyst to sRNase A. In the presence of substoichiometric amounts of wild-type PDI (360 nM), sRNase A (3.6 μM) is converted to active RNase A within approximately 2 h. The yield of native protein is, on average, only 50%. Data from triplicate reactions were fitted to an exponential rate curve and corrected for the uncatalyzed reaction to give a specific activity of 360 ± 69 nmol sRNase reactivated $min^{-1}\ \mu mol^{-1}$ (Table 2.1). Isomerization of sRNase A by CGHA:CGHA PDI was much less efficient than by wild-type with a specific activity of 28 ± 18 nmol sRNase reactivated $min^{-1}\ \mu mol^{-1}$. SGHC:SGHC PDI had no activity over the glutathione background (specific activity = 0.53 ± 0.83 nmol $min^{-1}\ \mu mol^{-1}$).

2.5 Discussion

Despite decades of research, our understanding of the mechanism by which PDI catalyzes disulfide bond isomerization is limited. Many questions remain to be answered regarding the individual catalytic steps of the isomerization reaction. The role of the *N*-terminal active-site thiolate in the initial attack of a nonnative substrate disulfide has been well established (Hawkins & Freedman, 1991; Darby *et al.*, 1994). Yet, it is still unclear how PDI initially recognizes its substrate. The *C*-terminal active-site cysteine residue appears to be involved in the rescue of trapped mixed-disulfide intermediates during catalysis of sRNase A isomerization (Walker & Gilbert, 1997). Yet, this residue is not required for complementation of *pdi1*Δ *S. cerevisiae* (Laboissière *et al.*, 1995b). More detailed chemical analyses of the mechanism of isomerization are required for a complete understanding of the pathway of native disulfide bond formation. The use of complex, nonhomogenous protein substrates has hindered kinetic study of the individual steps of catalysis.

We have addressed the need for a simplified assay for disulfide bond isomerization in the current study. By adapting the sequence of tachyplesin I to incorporate donor and acceptor fluorophores, we were able to use FRET to follow the rearrangement from nonnative to native disulfide bonds directly. The Trp–dansyl pair is an ideal choice for this assay. The small Förster radius allows us to observe the small distance changes that accompany the conversion of substrate to product. Preliminary investigation of alternative FRET donor–acceptor pairs did not give adequate sensitivity (data not shown).

The novel disulfide bond isomerization assay described here has a number of advantages over the most common assay, in which sRNase A is the substrate. Dns-sTI is a well defined substrate unlike sRNase A, which is a random mixture of nonnative protein conformations with batch-to-batch variability and thousands of potential mixed-disulfide intermediates with glutathione. The complexity of the sRNase A substrate means that some folding intermediates will be unproductive, even in the presence of PDI. Thus, in a typical sRNase A assay only 50% of the substrate is converted to native protein. By using a small peptide substrate, conformational traps are avoided and dns-sTI is converted completely to dns-nTI, simplifying kinetic analyses. The isomerization of sRNase by wild-type PDI takes 2 h to run to completion and is monitored indirectly by the appearance of ribonucleolytic activity. The isomerization of dns-sTI, on the other hand, is complete within 7 min and is monitored directly by FRET. Finally, the dns-sTI assay requires a small amount of enzyme (20 nM) for accurate and sensitive kinetic measurements while the sRNase A assay usually employs significantly higher concentrations of PDI (≥ 360 nM).

In the design of a disulfide bond isomerization assay, it is important that PDI catalyzes rearrangement of nonnative disulfides in the substrate peptide but does not reduce native disulfides in the product peptide. Accordingly, a challenge in the search for an ideal substrate is to identify a peptide with an appropriate reduction potential for isomerization by PDI. The reduction potential determined here of –0.313 V confirms that the disulfide bonds of dns-nTI are poor substrates for reduction by PDI ($E^{\circ\prime} = -0.180$ V), suggesting that PDI will recognize dns-nTI as a native conformation. Given that we observe only one E° value and that our FRET assay measures only a change in distance between *N*- and *C*-termini, it is possible that the E° value observed here is for the reduction of only one disulfide which destabilizes the β-sheet structure. The reduction potential of –0.313 V indicates that dns-nTI is more oxidizing than DTT ($E^{\circ} = -0.330$ V; (Cleland, 1964)) but more reducing than thioredoxin ($E^{\circ} = -0.270$ V; (Krause *et al.*, 1991)), the ubiquitous cellular reductant.

The isomerization of dns-sTI is mediated by a glutathione redox buffer alone. The first seconds of the reaction display unusual behavior for an isomerization assay, with a lag in product formation preceding the steady-state phase. The length of the lag decreases with increasing concentrations of glutathione (data not shown), and is approximately 20 s in the assay conditions reported here (120 μM total glutathione). A similar lag is observed in oxidative folding assays that use reduced RNase A as substrate and has been attributed to a slow initial oxidation step in which nonnative disulfide intermediates accumulate (Lyles & Gilbert, 1991; Walker *et al.*, 1996). In the dns-sTI assay, the lag could represent slow formation of mixed disulfide intermediates between glutathione and substrate peptide. Thiolate attack of reduced glutathione on substrate disulfides is inefficient at pH 7.6 due to a thiol pK_a value of 8.7 (Reuben & Bruice, 1976). In the presence of enzyme, the lag phase is not observed because mixed disulfide formation by PDI is significantly more efficient due to the depressed active-site thiol pK_a value (6.7; (Hawkins & Freedman, 1991)).

We have used the dns-sTI substrate to investigate the role of the more *C*-terminal cysteine residue of the PDI active site in the mechanism of isomerization. Previous studies have revealed that variants of PDI in which the *C*-terminal cysteine is replaced with serine or alanine, are less effective at refolding sRNase A (Lyles & Gilbert, 1994; Walker *et al.*, 1996; Walker & Gilbert, 1997). This finding has been confirmed in the current study. CGHA:CGHA PDI displays a specific activity that is only 7.7% that of wild-type PDI for isomerization of sRNase A (Table 2.1). As predicted, mutation of the *N*-terminal cysteine residue to alanine destroys all isomerase activity. The relative isomerase activity of CGHA:CGHA PDI in this study is slightly lower than values obtained in other studies using sRNase A as substrate (20–30% that of wild-type activity; (Walker *et al.*, 1996; Walker & Gilbert, 1997)). The discrepancy demonstrates the variability inherent to the sRNase assay. Different preparations of substrate and variations in assay method are difficult to avoid. Yet, our data support the hypothesis that the *C*-terminal cysteine residue is important for the rescue of trapped mixed-disulfide intermediates that accumulate during refolding of a complex substrate protein.

The isomerization of dns-sTI by CGHA:CGHA PDI is also less efficient than that catalyzed by the wild-type enzyme. For this substrate, however, the variant retains 50% of the activity of wild-type PDI. The replacement of the *C*-terminal cysteine residue with alanine is known to diminish the reactivity of the *N*-terminal cysteine residues. Previous studies showed that a CGHA variant of the a domain of human PDI displays a two-fold lower reactivity towards oxidized glutathione at pH 7.4 (Darby & Creighton, 1995a), suggesting that removal of the *C*-terminal active-site cysteine increases the pK_a of the *N*-terminal cysteine residue. The 50% decrease in the rate of isomerization of dns-sTI by CGHA:CGHA PDI could be explained simply by a decrease in the reactivity of the *N*-terminal cysteine residue. This result suggests that the refolding of a simple peptide substrate involves fewer trapped intermediates and therefore is less dependent on the escape pathway provided by the *C*-terminal cysteine residues (Figure 2.1).

The work we have described provides a new tool for studying the catalytic mechanism of disulfide bond isomerization. Unlike traditional isomerization assays, the peptide assay involves a minimal number of reaction intermediates and can be monitored in a continuous manner. Under the conditions used here, peptide substrate is completely converted to product in less than 10 min, allowing for rapid analysis. Additionally, the sensitivity of the assay requires relatively small amounts of catalyst for efficient turnover. The assay will be useful for future mechanistic studies of PDI and its homologs and should be a rapid method for the characterization of new small-molecule mimics of PDI.

Kinetic analysis using the dns-sTI substrate and a CGHA:CGHA variant of PDI have increased our understanding of the catalytic mechanism of isomerization. The rate of isomerization of sRNase A is significantly lower in the absence of the *C*-terminal active-site cysteine residue. The rate of dns-sTI isomerization, on the other hand, is only moderately decreased by the same alteration, and could be explained by an increase in the *N*-terminal cysteine pK_a value. These results suggest that with less complex substrates the escape mechanism for resolving trapped mixed-disulfide intermediates does not contribute significantly to catalysis.

Table 2.1 Catalysis of disulfide bond isomerization by PDI

Active sites	k_{cat}/K_m for dns-sTI (x 10^5 $M^{-1}s^{-1}$)	k_{cat}/K_m for dns-sTI (%)	Specific activity for sRNase A (U/µmol)[a]	Specific activity for sRNase A (%)
CGHC:CGHC	1.7 ± 0.52	100	360 ± 69	100
CGHA:CGHA	0.86 ± 0.35	50	28 ± 18	7.7
SGHC:SGHC	n.d.	n.d.	0.56 ± 0.83	0.16

n.d. not determined

[a] One unit (U) catalyzes the activation of 1 nmol sRNase A per minute in 100 mM Tris–HCl buffer (pH 7.6) containing EDTA (1.0 mM), GSH (1.0 mM), and GSSG (0.2 mM).

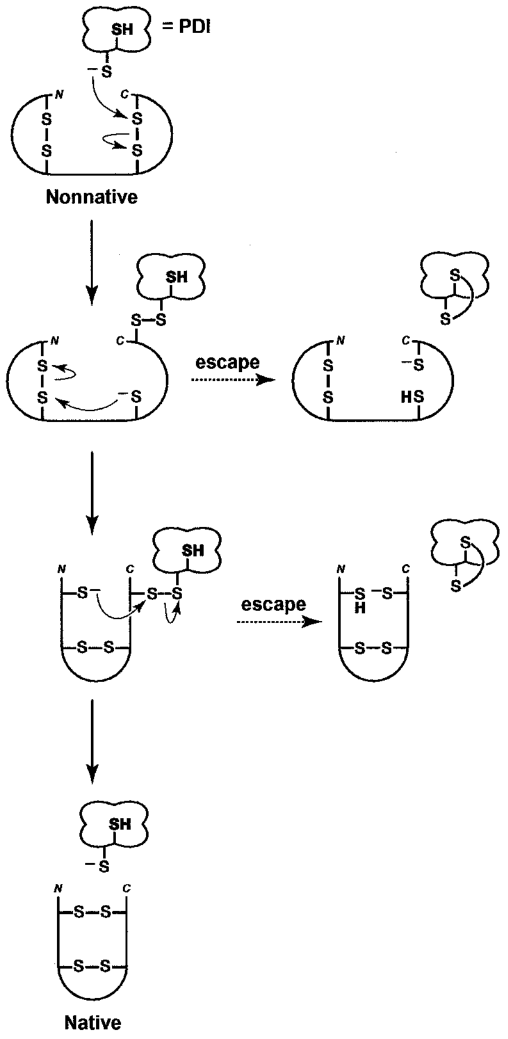


Figure 2.1 The mechanism of PDI-catalyzed isomerization is initiated by nucleophilic attack of the reactive *N*-terminal thiolate on a nonnative substrate disulfide bond. Further thiol–disulfide shuffling within the peptide substrate releases peptide substrate with a native configuration of disulfide bonds. Alternatively, the *C*-terminal active-site cysteine residue initiates reduction of trapped mixed disulfide intermediates. This escape mechanism involves cycles of reduction and reoxidation that eventually produce native peptide. (Reproduced from (Woycechowsky *et al.*, 1999)).

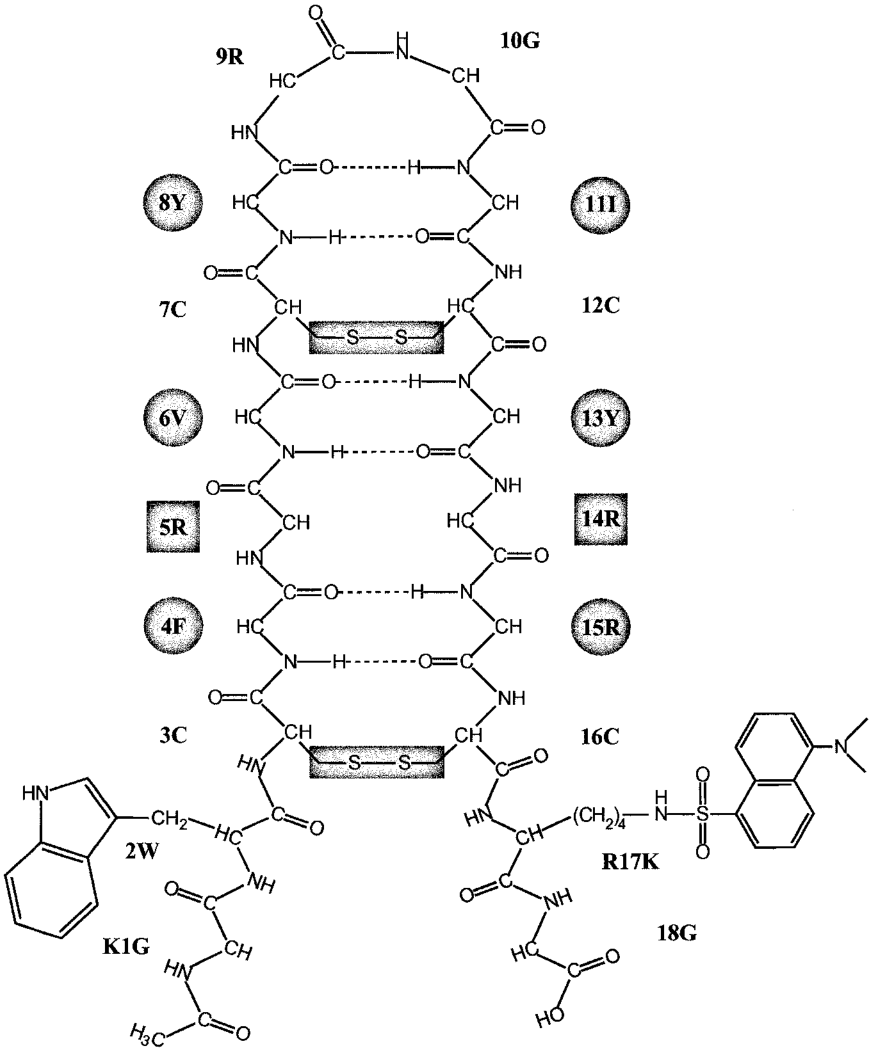


Figure 2.2 Tachyplesin I (TI) folds into a β-sheet stabilized by six hydrogen bonds and two disulfide bonds. Side chains indicated with circles lie above the plane of the β-sheet; side chains with rectangles lie below the plane of the β-sheet. For use in the disulfide bond isomerization assay, several substitutions to the native TI sequence were made. Lys1 was changed to a glycine residue, and Arg17 was changed to a lysine residue. The dansyl acceptor fluorophore was attached to the ε-amino group of Lys17. Trp2 is the donor fluorophore.

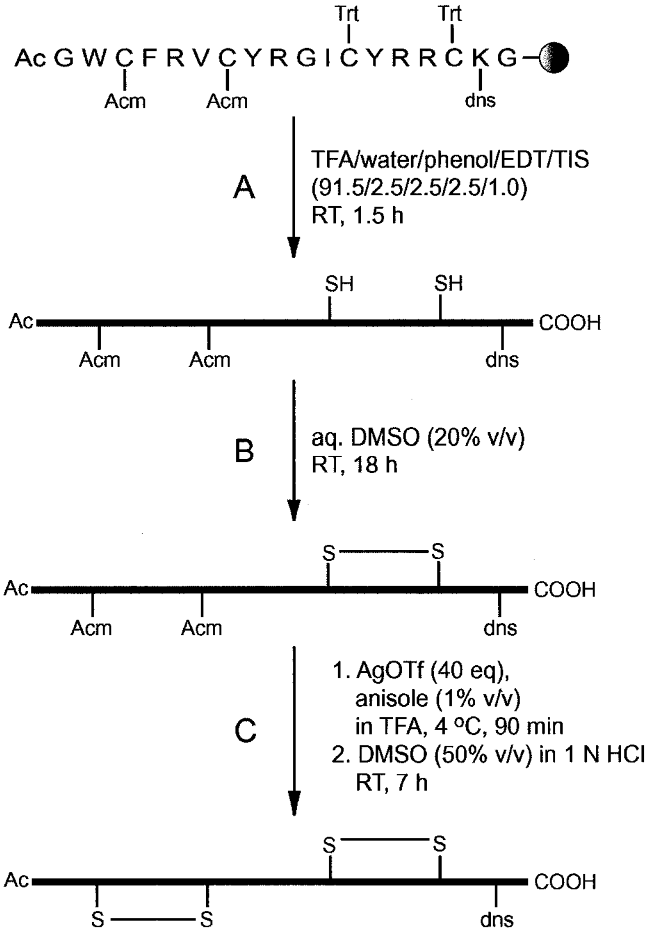

Trt
Trt
Ac G W C F R V C Y R G I C Y R R C K G
Acm
Acm
dns
A
TFA/water/phenol/EDT/TIS
(91.5/2.5/2.5/2.5/1.0)
RT, 1.5 h
SH
SH
Ac
COOH
Acm
Acm
dns
B
aq. DMSO (20% v/v)
RT, 18 h
S
S
Ac
COOH
Acm
Acm
dns
C
1. AgOTf (40 eq), anisole (1% v/v) in TFA, 4 °C, 90 min
2. DMSO (50% v/v) in 1 N HCl RT, 7 h
S
S
Ac
COOH
S
S
dns

Figure 2.3 Scheme for the synthesis of dns-sTI. The peptide was synthesized on solid-phase using standard Fmoc-protection methods. The lysine residue was protected with a dansyl (dns) group. The *N*-terminal cysteine residues were protected with acetamidomethyl (Acm) groups; the *C*-terminal cysteines were protected with trityl (Trt) groups. In step A, the peptide was cleaved from the resin, and all except the Acm protecting groups removed. Following HPLC purification, the first disulfide bond was formed in step B. Finally, the Acm groups were removed and the second disulfide bond formed in step C.

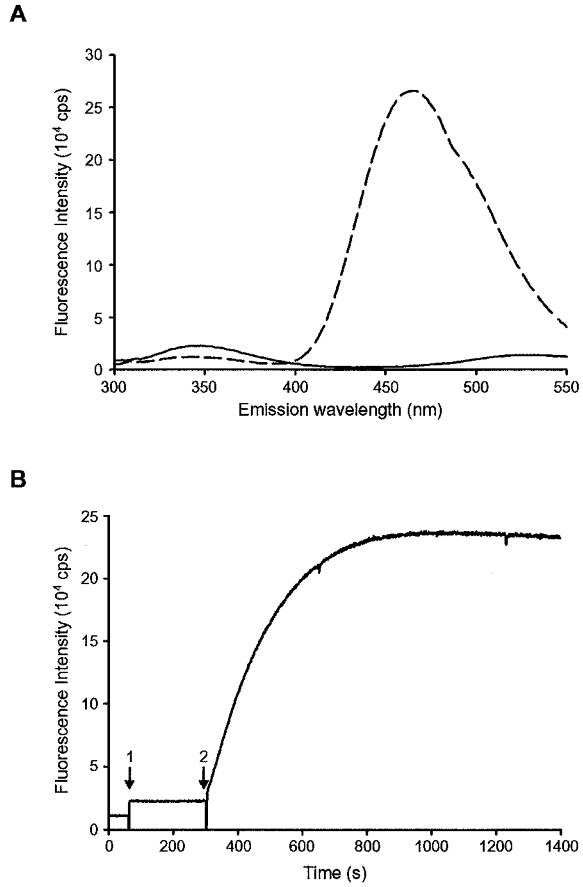


Figure 2.4 Fluoresence properties of dns-nTI and dns-sTI. **A**. Fluorescence emission intensity of dns-nTI (1.1 μM, dashed line) and dns-sTI (1.1 μM, solid line) as a function of wavelength upon excitation at 280 nm. Emission intensity was corrected for buffer (100 mM Tris–HCl buffer, pH 7.6, containing 1 mM EDTA). **B**. Fluorescence emission intensity (emission 465 nm, excitation 280 nm) of dns-sTI (1.1 μM) after addition of PDI (20 nM) and glutathione (116 μM GSH and 4 μM GSSG). Reaction was performed in 100 mM Tris–HCl buffer, pH 7.6, containing EDTA (1 mM) and IGEPAL CA-630 (0.01% w/v). At arrow 1, dns-sTI substrate was added; at arrow 2, PDI and glutathione were added.

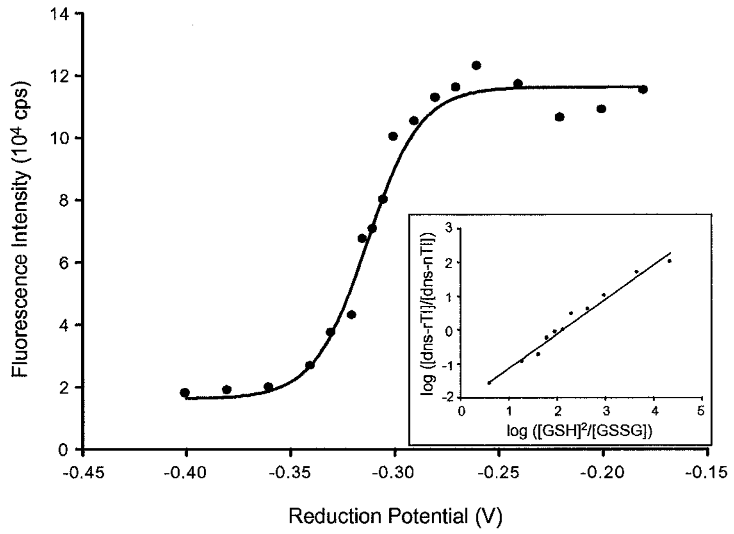


Figure 2.5 Fluorescence emission intensity of dns-nTI as a function of reduction potential. Thiol–disulfide exchange equilibria were established between dns-nTI (475 nM) and varying ratios of reduced to oxidized glutathione for 20 min before fluorescence measurements were obtained. The reduction potential of dns-nTI (–0.313 V) was obtained by a nonlinear least-squares analysis of the data according to the Nernst equation. Linear least-squares analysis of the Hill plot data (inset) indicated no cooperativity ($h = 1.018$) and $E^0 = -0.314$ V.

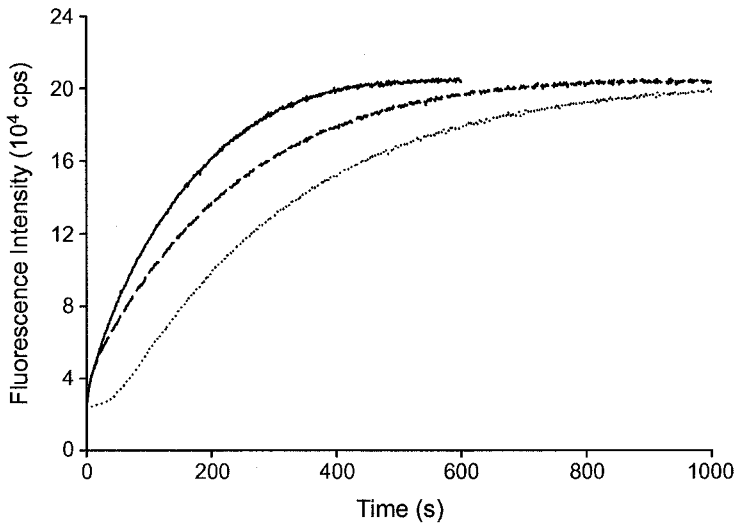


Figure 2.6 Isomerization of dns-sTI by PDI. Fluorescence emission intensity (emission 465 nm, excitation 280 nm) of dns-sTI (1.1 μM) was monitored after addition of wild-type PDI (20 nM, solid line), CGHA:CGHA PDI (20 nM, dashed line), or glutathione alone (116 μM GSH and 4 μM GSSG, dotted line). Reactions were performed in 100 mM Tris–HCl buffer (pH 7.6) containing EDTA (1 mM) and IGEPAL CA-630 (0.01% w/v). Reactions with PDI also contained glutathione (116 μM GSH and 4 μM GSSG).

Chapter Three

Production of Human Prolyl 4-Hydroxylase in *Escherichia coli*

This chapter was published as:

Kersteen, E. A., Higgin, J. J., and Raines, R. T. (2004). Production of human prolyl 4-hydroxylase in *Escherichia coli*. *Prot. Exp. Purif.* **38** 279-291.

3.1 Abstract

Prolyl 4-hydroxylase (P4H) catalyzes the post-translational hydroxylation of proline residues in collagen strands. The enzyme is an $\alpha_2\beta_2$ tetramer in which the α subunits contain the catalytic active sites and the β subunits (protein disulfide isomerase) maintain the α subunits in a soluble and active conformation. Heterologous production of the native $\alpha_2\beta_2$ tetramer is challenging and had not been reported previously in a prokaryotic system. Here, we describe the production of active human P4H tetramer in *Escherichia coli* from a single bicistronic vector. P4H production requires the relatively oxidizing cytosol of Origami™ B(DE3) cells. Induction of the wild-type α(I) cDNA in these cells leads to the production of a truncated α subunit (residues 235–534), which assembles with the β subunit. This truncated P4H is an active enzyme, but has a high K_M value for long substrates. Replacing the Met235 codon with one for leucine removes an alternative start codon and enables production of full-length α subunit and assembly of the native $\alpha_2\beta_2$ tetramer in *E. coli* cells to yield 2 mg of purified P4H per L of culture (0.2 mg per g of cell paste). We also report a direct, automated assay of proline hydroxylation using high-performance liquid chromatography. We anticipate that these advances will facilitate structure–function analyses of P4H.

3.2 Introduction

Collagen, the most abundant protein in animals, is a right-handed triple helix composed of three polyproline type-II strands. Each strand contains a repeating sequence of Xaa-Yaa-Gly, where the Xaa residue is often (2*S*)-proline (Pro) and the Yaa residue is often (2*S*,4*R*)-4-hydroxyproline (Hyp). The Hyp residues in collagen strands arise from the post-translational modification of Pro residues by prolyl 4-hydroxylase (P4H). This hydroxylation reaction is critical for the folding of the collagen triple helix (Chopra & Ananthanarayanan, 1982; Bulleid *et al.*, 1996; Snellman *et al.*, 2000). Moreover, the conformational stability of a collagen triple helix correlates with its Hyp content (Berg & Prockop, 1973c; Burjanadze, 2000). P4H is known to be an essential enzyme for the nematode *Caenorhabitis elegans* (Friedman *et al.*, 2000; Winter & Page, 2000).

P4H resides in the lumen of the endoplasmic reticulum (ER), where the folding and assembly of procollagen occurs. In vertebrates, P4H is found as an $\alpha_2\beta_2$ tetramer; in *C. elegans*, P4H can also exist as a $\alpha\beta$ dimer (Myllyharju & Kivirikko, 1997). Each α subunit contains a catalytic active site for proline hydroxylation (Lamberg *et al.*, 1995). The β subunit is protein disulfide isomerase (PDI) (Koivu *et al.*, 1987b), a thiol–disulfide oxidoreductase with numerous roles in the cell (Gilbert, 1997; Guzman, 1998; Kersteen & Raines, 2003). The primary roles of PDI in the P4H tetramer appear to be to keep the α subunit in a soluble and active form (Vuori *et al.*, 1992b; Vuori *et al.*, 1992c) and to retain the enzyme in the ER through the KDEL sequence at its *C* terminus (Kivirikko *et al.*, 1989). The active-site cysteine residues of PDI are not required for P4H activity or tetramer assembly (Vuori *et al.*, 1992c). Three human isoforms of the α subunit have been identified: α(I), α(II), and α(III) (Helaakoski *et al.*, 1989; Annunen *et al.*, 1997;

Kukkola *et al.*, 2003). P4Hα(I) accounts for most of the prolyl 4-hydroxylase activity in a variety of cell types (Annunen *et al.*, 1997; Kivirikko & Myllyharju, 1998; Kukkola *et al.*, 2003).

P4H is a member of a class of α-ketoglutarate-dependent non-heme Fe(II) dioxygenase enzymes (Fox, 1998; Guzman, 1998; Kivirikko & Pihlajaniemi, 1998). These enzymes require α-ketoglutarate, Fe(II), and O_2 for catalysis. During proline hydroxylation (Figure 3.1), α-ketoglutarate is decarboxylated oxidatively to form succinate (Rhoads & Udenfriend, 1968). One atom of O_2 is incorporated into the nascent hydroxyl group of Hyp, and the other into succinate (Cardinale *et al.*, 1971). Ascorbate is required to rescue inactivated enzyme that accumulates from decarboxylation of α-ketoglutarate that is uncoupled from substrate hydroxylation (Myllylä *et al.*, 1978; de Jong *et al.*, 1982; Myllylä *et al.*, 1984).

α Subunit residues that are critical for catalysis have been identified by site-directed mutagenesis. The iron-binding ligands are His412, Asp414, and His483 (Lamberg *et al.*, 1995; Myllyharju & Kivirikko, 1997). Lys493 binds to the C-5 carboxyl group of α-ketoglutarate, and His501 likely binds to its C-1 carboxyl group (Myllyharju & Kivirikko, 1997). Additional insights into the catalytic mechanism have been gained by comparison of P4H to other α-ketoglutarate-dependent dioxygenases (Schofield & Zhang, 1999; Costas *et al.*, 2004). Substrate binding has been investigated with an assay that measures hydroxylation of (Pro–Pro–Gly)$_n$ collagen mimics indirectly—by monitoring the release of $[^{14}C]CO_2$ during decarboxylation of $[1\text{-}^{14}C]$α-ketoglutarate (Rhoads & Udenfriend, 1968; Kivirikko & Myllylä, 1982). Increasing the length of collagen-like substrates leads to tighter binding in the enzymic active site (Kivirikko *et al.*, 1972). The lack of a known three-dimensional structure for P4H, however, has made it difficult to further decipher details of the mechanism of hydroxylation. The difficulties faced in the assembly of recombinant P4H tetramer have significantly hindered crystallographic study of the enzyme.

Production of the P4H tetramer is not trivial. Attempts to assemble an active P4H tetramer from its individual α and β subunits *in vitro* have been unsuccessful (Nietfeld *et al.*, 1981; Koivu & Myllylä, 1986). Although the P4H subunits dissociate easily, reassociation of the α subunit (which is insoluble) with PDI to form an active tetramer has not been possible. It has been suggested (John & Bulleid, 1996) that *in vivo* assembly of P4H requires chaperone proteins, present in the ER of eukaryotic cells, to maintain the α subunit in a soluble form until PDI can bind and allow formation of native tetramer.

Human P4H has been produced heterologously in baculovirus (Vuori *et al.*, 1992b). Baculovirus expression vectors encoding the α and β subunits were used to infect *Spodoptera frugiperda* (Sf9) insect cells. Association of the subunits to form active tetramer (5 mg per L of culture) was observed upon cotransfection. Recently, assembly of active human P4H tetramer was also achieved in the yeast *Pichia pastoris* (Vuorela *et al.*, 1997). Production of the α subunit along with the β subunit, attached to the *Saccharomyces cerevisiae* α mating factor pre-pro signal sequence, yields active P4H tetramer in yeast. Higher levels of P4H tetramer (15 mg per L of culture) were obtained by coexpressing the α and β subunits with recombinant human type III procollagen in yeast (Vuorela *et al.*, 1997).

We suspected that developing an *Escherichia coli* expression system for P4H could be of substantial benefit to its further study. Heterologous production of recombinant proteins has been successful in a number of host organisms (Higgins & Hames, 1999; Chu & Robinson, 2001; Andersen & Krummen, 2002). Still, *E. coli* remains the most powerful (Baneyx, 1999; Swartz, 2001). For example, almost all proteins that have been crystallized (97%) and have had their structures determined by X-ray diffraction analysis (93%) were produced in *E. coli* (Goulding & Perry, 2003). Moreover, the generation of protein variants is comparatively facile in *E. coli* systems (Sambrook & Russell, 2001).

Here, we report the production of active human P4H in *E. coli*. The two P4H subunits were expressed from a bicistronic vector in cells that have an oxidizing cytosol (Origami™ B(DE3)). In addition to intact P4H, we also inadvertently produced a truncated variant of P4H (P4Hα(I)235–534/β) by the initiation of translation from an internal methionine residue. Finally, we report the development of a facile assay for proline hydroxylation. These advances could facilitate structure–function analyses of P4H.

3.3 Materials and Methods

Materials. E. coli strains BL21(DE3), Origami™ B(DE3), and Rosetta-gami™ 2(DE3), and the pET22b(+) expression vector were from Novagen (Madison, WI). Enzymes for DNA manipulation were from Promega (Madison, WI). DNA oligonucleotides for mutagenesis and sequencing were from Integrated DNA Technologies (Coralville, IA). DNA sequencing reactions were performed with the BigDye kit from Applied Biosystems (Foster City, CA) and CleanSeq magnetic beads from Agencourt Bioscience (Beverly, MA). DNA sequences were determined by using capillary arrays on an Applied Bioscience automated sequencing instrument at the University of Wisconsin Biotechnology Center.

Poly(proline) was from Sigma Chemical (St. Louis, MO). Prosieve protein markers for SDS–PAGE and PAGEr Gold Precast gels (7.5% Tris–glycine) for native PAGE were from Cambrex Bio Science Rockland (Rockland, ME). Native PAGE molecular-mass standards were from

Amersham Biosciences (Piscataway, NJ). Prestained SDS–PAGE molecular-mass standards were from Bio-Rad (Hercules, CA). (Pro–Pro–Gly)$_{10}$ was from Peptides International (Louisville, KY). [1-^{14}C]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.

Luria–Bertani (LB) medium contained (in 1.0 L) tryptone (10 g), yeast extract (5 g) and NaCl (10 g). Terrific broth (TB) medium contained (in 1.0 L) tryptone (12 g), yeast extract (24 g), K_2HPO_4 (72 mM), and KH_2PO_4 (17 mM), and glycerol (4 mL). All media were prepared in deionized, distilled water and autoclaved.

Human P4H produced in a baculovirus system (Vuori *et al.*, 1992a) was a generous gift of Johanna Myllyharju (University of Oulu) and was used as a standard for the experiments reported herein.

Instrumentation. UV absorbance measurements were made with a Cary 50 spectrophotometer from Varian (Palo Alto, CA). High performance liquid chromatography (HPLC) was performed with a system from Waters (Milford, MA) that includes two 515 pumps, a 717plus autosampler, and a 996 photodiode array detector. The system was controlled with the manufacturer's Millennium32 (Version 3.20) software. Fast performance liquid chromatography (FPLC) was performed with an ÄKTA system from Amersham Pharmacia (Piscataway, NJ) and analyzed with the UNICORN Control System. Scintillation counting was performed on a Wallac 1450 MicroBeta TriLux liquid scintillation counter from PerkinElmer (Wellesley, MA).

Design of P4H Expression Vector. Our strategy for expressing active P4H tetramer in *E. coli* was to clone cDNAs encoding the α and β subunit into the same plasmid. From this bicistronic vector, transcription of both cDNAs could be accomplished from the same T7 promoter, and each subunit would have its own ribosome binding site (rbs) for the initiation of translation. cDNA encoding human PDI was a kind gift from Alan D. Attie (University of Wisconsin–Madison). That cDNA, without its signal sequence encoding region, was inserted between the *Nde*I and *Bam*HI restriction sites of a pET22b(+) expression vector to give plasmid pBK1.PDI1.

cDNA encoding P4Hα(I) (PA11 clone), which had been isolated from HeLa cells and inserted into a pBSKS– vector (pBS.LF17-1), was a kind gift from Lisa Friedman (University of Wisconsin–Madison). DNA encoding P4Hα(I) was isolated from pBS.LF17-1 by the PCR using primers that flank regions on the 5′ side of the translation initiation codon and the 3′ side of the stop codon, each including a *Bam*HI restriction site. The PCR fragment was cloned into the TOPO vector (Invitrogen, Carlsbad, CA) and then digested with *Bam*HI. The resulting fragment was then ligated into pBK1.PDI1 that had been digested with *Bam*HI, yielding plasmid pBK1.PDI1.P4H1.

DNA encoding the signal sequence of P4Hα(I) was removed in a series of steps. QuikChange (Stratagene, La Jolla, CA) site-directed mutagenesis was used to: (1) add an rbs approximately 15 bp on the 5′ side of the first encoded amino acid of the processed protein, (2) replace the last codon for the signal sequence region with an ATG start codon, and (3) add an *Nde*I site immediately on the 5′ side of the engineered start codon. The resulting plasmid pBK1.PDI1.P4H5 was used to produce the P4Hα(I)235–534/β enzyme, which is a P4H oligomer with an α subunit of only 32 kDa.

A plasmid encoding the P4Hα(I) subunit alone was produced by digesting pBK1.PDI1.P4H5 with *Nde*I, removing the DNA fragment encoding PDI, and ligating the vector. The resulting construct (pBK1.P4H5) was subjected to QuikChange mutagenesis to add a *Bam*HI site on the 5′ side of the pET22b(+) rbs, yielding plasmid pBK1.P4H6. *Bam*HI digestion of pBK1.PDI1 and pBK1.P4H6, followed by ligation, resulted in a vector with the PDI cDNA preceding the P4Hα(I) cDNA, both having the polylinker of pET22b(+) on the 5′ side of their start codons (pBK1.PDI1.P4H6). Finally, the ATG codon of Met235 of the α subunit was replaced with the CTT codon of leucine by QuikChange mutagenesis to yield plasmid pBK1.PDI1.P4H7 (Figure 3.2), which was used to produce the full-length P4H tetramer.

Production of P4H. Expression vectors were transformed into BL21(DE3) or Origami™ B(DE3) cells by electroporation. Transformations in BL21(DE3) were grown on LB agar containing ampicillin (100 μg/mL); transformations in Origami™ B(DE3) were grown on LB agar containing ampicillin (100 μg/mL), kanamycin (15 μg/mL), and tetracycline (12.5 μg/mL). A starter culture in LB medium (25 mL), supplemented with antibiotics as above, was inoculated with a fresh colony within 24 h of transformation. Cells were grown overnight for 10–12 h, and then harvested by centrifugation at 5000g for 10 min. The resulting cell pellet was resuspended in fresh LB medium (25 mL) and used to inoculate to $OD_{600} = 0.01$ either a 1 L culture of TB medium containing ampicillin (200 μg/mL) for the growth of BL21(DE3) cells, or TB medium containing ampicillin (200 μg/mL), kanamycin (15 μg/mL), and tetracycline (12.5 μg/mL) for the growth of Origami™ B(DE3) cells. The inoculated culture was shaken (210 rpm) at 37 °C until reaching $OD_{600} = 1.7–1.8$. cDNA expression was then induced by the addition of isopropyl-

1-thio-β-D-galactopyranoside (IPTG, 500 μM) and shaking at 37 °C for 4 h in BL21(DE3) cells or at 23 °C for 12–18 h in Origami™ B(DE3) cells. Cells were harvested by centrifugation and resuspended in 100 mM glycine, 10 mM Tris, 100 mM NaCl, pH 7.8. Approximately 10 g of cell paste was obtained from each L of culture. Harvested and resuspended cells (10 g of cell paste in 20 mL) were lysed by passage three times through a French pressure cell. Insoluble material was removed by centrifugation at 50,000g for 1 h at 4 °C in an Optima™ XL-80K ultracentrifuge from Beckman Coulter (Fullerton, CA). Both subunits of P4H remain in the soluble lysate fraction.

Purification of P4H. Purification of P4H essentially follows procedures published previously (Kivirikko & Myllylä, 1987). The lysate was first purified by ammonium sulfate fractionation. Saturated ammonium sulfate was added to the lysate to 30% saturation, and the pellet was removed by centrifugation at 15,000g for 30 min. The supernatant was precipitated by adding ammonium sulfate to 65% saturation. The pellet was collected by centrifugation at 15,000g for 30 min and resuspended in 100 mM glycine, 10 mM Tris, 100 mM NaCl, pH 7.8. The 30–65% ammonium sulfate fraction was dialyzed overnight against 4 L of the same buffer.

Poly(proline) resin was prepared essentially as described previously (Tuderman *et al.*, 1975; Lindberg *et al.*, 1988). Cyanogen bromide-activated Sepharose 4 resin (Sigma Chemical) was swelled and rinsed with 1 mM HCl (700 mL) and then equilibrated with coupling buffer (0.1 M $NaHCO_3$, 0.1 M NaCl, pH 8.3, 800 mL). Poly(L-proline) (average molecular mass: 15–28 kDa) was suspended in 75 mL coupling buffer and heated gently to help break up insoluble material to form an "opalescent suspension". The poly(L-proline) solution was cooled with stirring at 4 °C and then coupled to equilibrated resin at 4 °C overnight with stirring. After letting the resin settle, buffer and insoluble, uncoupled poly(L-proline) were decanted and the resin was washed with coupling buffer (700 mL). Finally, the resin was washed with 100 mM Tris–HCl, pH 8.0 (500 mL). The column was equilibrated with 100 mM glycine, 10 mM Tris, 100 mM NaCl, pH 7.8 (500 mL). Following protein purification, the poly(L-proline) affinity column was washed with urea (6 M, 500 mL) to remove any remaining protein.

The ammonium sulfate fraction was diluted to 10 mg/mL (total volume: 180 mL) and loaded onto the column (7 cm × 7 cm^2) of poly(proline) resin at a flow rate of 1 mL/min. The column was washed with 100 mM glycine, 10 mM Tris, 100 mM NaCl, pH 7.8 until the OD_{230} and OD_{280} of the eluate were <0.05 (approximately 250 mL). P4H was eluted with 20 mL of 5 mg/mL poly(proline) (average molecular mass: 7 kDa) in the same buffer, followed by buffer alone. The absorbance at 230 and 280 nm of each fraction (4 mL) was measured, and appropriate fractions were pooled. Pooled fractions were concentrated with a Vivaspin concentrator (20 mL, molecular mass cutoff: 10 kDa; Vivascience AG, Hanover, Germany) and dialyzed against 25 mM sodium phosphate, 10 mM glycine, 50 mM NaCl, pH 7.8 in preparation for anion-exchange chromatography.

Cation-exchange chromatography was used as an alternative purification step for the truncated P4H, which did not bind to the poly(proline) resin. The 30–65% ammonium sulfate fraction was dialyzed against 25 mM sodium phosphate, 10 mM glycine, 50 mM NaCl, pH 7.0 and then loaded onto a column (20 mL) of Mono S cation-exchange resin (HiLoad 16.60, Amersham Biosciences) that had been equilibrated with the same buffer. The anionic P4H protein did not bind to the resin, and was washed through with 60 mL of buffer. P4H was concentrated with a Vivaspin concentrator and dialyzed against 25 mM sodium phosphate, 10 mM glycine, 50 mM NaCl, pH 7.8 in preparation for anion-exchange chromatography.

In the next purification step, dialyzed protein from affinity chromatography (or cation-exchange chromatography) was injected at 1 mL/min onto a column (8 mL) of Resource Q anion-exchange resin (Amersham Biosciences) that had been equilibrated with 25 mM sodium phosphate, 10 mM glycine, 50 mM NaCl, pH 7.8. After washing with equilibration buffer (16 mL), the column was eluted with a linear gradient (60 + 60 mL) of NaCl (50–430 mM) in equilibration buffer. Pooled fractions were concentrated with a Vivaspin concentrator and dialyzed overnight against 100 mM glycine, 10 mM Tris, 100 mM NaCl, pH 7.8.

In the final purification step, dialyzed protein from anion-exchange chromatography was loaded onto a column (318 mL) of Superdex-200 gel-filtration resin (Amersham Biosciences) that had been preequilibrated with 100 mM glycine, 10 mM Tris, 100 mM NaCl, pH 7.8. The column was eluted at a flow rate of 1.5 mL/min. Peak fractions were combined and concentrated. A correlation between protein molecular mass and column retention time was made by constructing a standard curve using a calibration kit (29–700 kDa) for gel-filtration chromatography (Sigma Chemical).

Analysis of P4H. During the purification of P4H, protein fractions were analyzed by SDS–PAGE in a gel containing 10% (w/v) poly(acrylamide) under reducing conditions, followed by staining with Coomassie Brilliant Blue R-250 or immunoblotting. Immunoblot analysis was performed with a polyclonal antibody to human PDI, which was a generous gift from Alan D. Attie, and anti-human P4Hα(I) monoclonal antibody (ICN Biomedicals, Costa Mesa, CA). Primary antibodies were detected with HRP-conjugated secondary antibodies and ECL Western Blotting Detection reagents (Amersham Biosciences).

Native PAGE was performed under non-reducing conditions to estimate the molecular mass of the truncated P4H. In this analysis, the truncated P4H and molecular-mass standards were subjected to electrophoresis on native gels containing 6, 8, 10, and 12% (w/v) poly(acrylamide).

In-gel trypsin digestion of proteins followed by MALDI–MS was performed at the University of Wisconsin Biotechnology Center. Protein bands excised from SDS–PAGE gels were destained and digested with trypsin. MALDI–MS was used to detect peptides isolated from the digest. The Mascot search engine from Matrix Science (www.matrixscience.com) was used to perform a peptide mass fingerprint search, allowing carbamidomethyl cysteine and oxidation of methionine as variable modifications, a peptide mass tolerance of ±1 kDa, and a maximum of 1 missed cleavage.

Measurement of Protein Concentrations. Total protein concentration was determined for crude samples with the Micro BCA™ Protein Assay Reagent Kit (Pierce Biotechnology, Rockford, IL) according to the manufacturer's instructions, except that samples were incubated at 37 °C for 2 h before taking absorbance measurements. Extinction coefficients were determined at 280 nm for P4H by using the method of von Hippel (Gill & von Hippel, 1989). Values obtained (ε_{280} = 290,000 $M^{-1}cm^{-1}$ for full-length P4H; ε_{280} = 190,000 $M^{-1}cm^{-1}$ for P4Hα(I)235–534/β) were used to calculate the P4H concentration in purified samples.

Synthesis of P4H Substrates. A fluorescent tetrapeptide substrate (dansyl-Gly–Phe–Pro–Gly–OEt) was synthesized by standard solution-phase methods. Boc-Phe–Pro-OH (870 mg, 2.4 mmol), glycine ethyl ester·HCl (341 mg, 2.4 mmol), dicyclohexylcarbodiimide (623 mg, 2.4 mmol), 1-hydroxybenzotriazole (330 mg, 2.4 mmol), and *N*-ethylmorpholine (309 μL, 2.4 mmol) were dissolved in tetrahydrofuran (10 mL), and the resulting solution was stirred at room temperature overnight. The product was extracted into CH_2Cl_2 (50 mL) from saturated aqueous $NaHCO_3$ (50 mL). The organic layer was washed with aqueous citric acid (50 mL, 5% w/v) and brine (50 mL), and concentrated by rotary evaporation. The Boc protecting group was removed by stirring the crude residue in CH_2Cl_2 (20 mL) with trifluoroacetic acid (7.3 mL) at 0 °C for 3 h. The intermediate product (Phe–Pro–Gly-OEt) was purified by silica gel chromatography, eluting with methanol (8% v/v) in CH_2Cl_2. Solvent was removed from the fractions containing the desired compound by rotary evaporation, and the residue was dissolved in dimethylformamide (20 mL). 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide (631 mg, 3.3 mmol), HOBt (349 mg, 2.6 mmol), *N*-ethylmorpholine (714 μL, 5.6 mmol), and dansylglycine (802 mg, 2.6 mmol) were added, and the reaction mixture was allowed to stir for 48 h. Water (10 mL) was added to quench the reaction. The product was extracted into ethylacetate (50 mL) and washed with aqueous HCl (5% v/v, 50 mL), saturated aqueous $NaHCO_3$ (50 mL), and brine (50 mL). The initial aqueous layer was extracted further into CH_2Cl_2 (50 mL), and the CH_2Cl_2 layer was washed with aqueous HCl (5% v/v, 50 mL), saturated aqueous $NaHCO_3$ (50 mL), and brine (50 mL). The organic layers were combined and dried over anhydrous $MgSO_4(s)$, and the solvent was removed by rotary evaporation. The product was purified by silica gel chromatography, eluting with methanol (4% v/v) in CH_2Cl_2 to yield dansyl-Gly–Phe–Pro–Gly-OEt as a bright yellow oil (328 mg, 21.4%). ESI–MS m/z: $[M + Na]^+$ calcd, 660.2468; found, 660.2438.

The corresponding product peptide (dansyl-Gly–Phe–Hyp–Gly-OEt) was used as an HPLC standard. The Hyp-containing peptide was synthesized by using standard Fmoc-protection strategies with HATU activation on an Applied Biosystems Pioneer automated synthesizer. Dansyl-Gly–Phe–Hyp–Gly-OH was cleaved from the solid support with trifluoroacetic acid

(1 mL) and dripped into ice-cold ethyl ether. The precipitated peptide was collected by centrifugation, and the crude product (341 mg, 0.5 mmol) was dried under vacuum. The dried peptide was heated at reflux with $SOCl_2$ (80 mL, 1.1 mmol) in ethanol (20 mL) for 2 h. The solvent was removed by rotary evaporation, and the product was purified by silica gel chromatography, eluting with methanol (6% v/v) in CH_2Cl_2 to yield dansyl-Gly–Phe–Hyp–Gly-OEt as a yellow oil (77 mg, 16.0%). ESI–MS *m/z*: $[M + Na]^+$ calcd, 676.2417; found, 676.2422.

Assays of Enzymatic Activity. An HPLC-based assay was developed to monitor hydroxylation of proline by P4H. Assays were performed at 30 °C in 100 μL of 50 mM Tris–HCl buffer (pH 7.8) containing bovine serum albumin (1 mg/mL), catalase (100 μg/mL), dithiothreitol (100 μM), $FeSO_4$ (50 μM), P4H, and α-ketoglutarate (500 μM). The tetrapeptide substrate (dansyl-Gly–Phe–Pro–Gly-OEt, 92 mM stock in methanol) was added to initiate the reaction. Aliquots (30 μL) were withdrawn at known times and quenched by boiling for 30 s. All assays were performed in triplicate.

Conversion to hydroxylated product (dansyl-Gly–Phe–Hyp–Gly-OEt) was monitored by HPLC using a reversed-phase analytical C18 column (4.6 × 250 mm) from Varian (Lake Forest, CA) equipped with an Adsorbosphere guard cartridge from Alltech (Deerfield, IL). Peptides were separated by isocratic elution with acetonitrile/water (1:1) at 1.0 mL/min and quantified by integration of the $A_{337.5\text{ nm}}$ according to a processing method developed with the Millennium32 software. Calibration curves were developed for the substrate and product peptides so as to correlate peak area with peptide concentration. The identity of HPLC peaks was confirmed by liquid chromatography/mass spectrometry (LC/MS) at the University of Wisconsin Biotechnology Center.

A more traditional P4H assay, which monitors the release of [^{14}C]CO_2 from [1-^{14}C]α-ketoglutarate, was used to confirm the validity of the new HPLC-based assay. Methods followed those described elsewhere (Rhoads & Udenfriend, 1968; Kivirikko & Myllylä, 1982). Briefly, bovine serum albumin (1 mg/mL), catalase (100 μg/mL), dithiothreitol (100 μM), ascorbate (2 mM), $FeSO_4$ (50 μM), P4H, and a substrate were combined (500 μL total reaction volume) on ice. Substrates used in this assay include the tetrapeptide described above as well as $(Pro\text{–}Pro\text{–}Gly)_{10}$. The latter substrate was boiled for 5 min and then placed on ice, immediately before performing the reaction. A hanging well that contained filter paper soaked with aqueous NaOH (2 M) was inserted through a rubber stopper that was used to seal each reaction. [1-^{14}C]α-Ketoglutarate (100 μM, 185000 dpm) was added to initiate the reaction. After 15 min at 30 °C, reactions were quenched by the addition of 500 μL of aqueous HCl (2 M) and allowed to sit at room temperature for 60 min. The filter paper was removed and dried under a bright light for 1.5 h. Liquid scintillation cocktail (3 mL, Wallac Optiphase SuperMix from PerkinElmer, Wellesley, MA) was added to each piece of paper, and ^{14}C was quantified by scintillation counting (5 min). All reactions were performed in duplicate and corrected for the rate of decarboxylation in the absence of the peptide substrate.

3.4 Results

Production of P4H from pBK1.PDI1.P4H5. The bicistronic plasmid pBK1.PDI1.P4H5 was constructed for the production of human P4H by placing DNA encoding the α(I) and β subunits into pET22b(+), under control of the same T7 promoter, and giving each cDNA its own rbs. Expression of cDNAs encoding both subunits was induced in BL21(DE3) cells as indicated by SDS–PAGE (data not shown), but the cell lysate from these cells had no detectable prolyl 4-

hydroxylase activity. In addition, no tetrameric protein could be detected by native PAGE (data not shown). Apparently, the assembly of active P4H tetramer was not achieved in BL21(DE3) cells. Addition of iron(II) sulfate, zinc(II) chloride, ascorbate, and/or α-ketoglutarate to the growth media did not promote assembly. Likewise, inclusion of denatured collagen or denatured (Pro–Pro–Gly)$_{10}$ in the lysis buffer did not assemble an active tetramer. Finally, attempts were made to assemble P4H tetramer *in vitro*. Refolding of insoluble P4Hα(I) expressed from pBK1.P4H5 with cell lysate containing PDI expressed from pBK1.PDI1 in the presence of glutathione, ascorbate, iron(II) sulfate, glycine, arginine, and/or denatured collagen was not successful.

Instead, pBK1.PDI1.P4H5 was induced in Origami™ B(DE3) cells at 23 °C. SDS–PAGE indicated that induction was successful (Figure 3.3A). In addition, prolyl 4-hydroxylase activity was detected in a crude cell lysate (data not shown).

Purification of Truncated P4H. SDS–PAGE of whole cell samples after induction of pBK1.PDI1.P4H5 in Origami™ B(DE3) cells revealed the induction of a 32-kDa protein, in addition to a protein (PDI) of 55 kDa (Figure 3.3A). Interestingly, induction of the 32-kDa protein was not observed in BL21(DE3) cells (data not shown). P4H in the 30–65% ammonium sulfate fraction of this cell lysate could not be purified with poly(proline)-affinity chromatography. Instead, the enzyme was purified by cation-exchange chromatography. Although P4H did not bind to a cation-exchange resin, many other proteins did (Figure 3.3A). Additional purification was achieved by anion-exchange chromatography, followed by gel-filtration chromatography. Following these two purification steps, SDS–PAGE revealed that >90% of the protein was in two bands corresponding to proteins of approximately 55 kDa and 32 kDa (Figure 3.3A). Approximately 2 mg of this P4H oligomer was obtained per L of growth medium.

The 32-kDa protein was digested with trypsin *in situ* after SDS–PAGE. The protein fragments indicated that the 32-kDa protein derived from the *C*-terminus of P4Hα(I). Specifically, ten peptides had masses corresponding to trypsin fragments of the human P4Hα(I) subunit between Gly260 and Glu534, which is the most *C*-terminal residue. This α subunit fragment is not likely to be an *E. coli* degradation product, as it is observed in whole cell extract immediately following induction with IPTG.

No induction of full-length α subunit was apparent. In-gel trypsin digestion and MALDI–MS of the purified 55-kDa band from SDS–PAGE revealed eight peptides from PDI but none from the α subunit. Immunoblot analysis of crude cell lysate and purified protein showed a strong response with a polyclonal PDI antibody, but no detectable response with a monoclonal antibody against human P4Hα(I) (Figure 3.3B and 3.3C).

A tetrameric P4H protein ($\alpha_2\beta_2$) containing PDI (55 kDa) and the *C*-terminal fragment of P4Hα(I) (32 kDa) would have a molecular mass of 174 kDa, whereas a dimeric P4H protein (αβ) would have a molecular mass of 87 kDa. The elution of the truncated P4H during analytical gel-filtration chromatography corresponds to a protein of 140 kDa, which is between the mass expected for dimer and tetramer. Full-length P4H eluted during analytical gel-filtration chromatography at 30 mL after the void volume, which corresponds to a protein of 380 kDa. Yet, the molecular mass of full-length P4H is 228 kDa. Thus, the structure of the wild-type tetramer is likely to be non-spherical. If the structure of the truncated P4H is also non-spherical, then the stoichiometry of its subunits cannot be determined by analytical gel-filtration chromatography.[1]

Native PAGE of the purified protein was performed with a range of poly(acrylamide) concentrations (6–12%) alongside a set of proteins of known molecular mass (Figure 3.4A). A calibration plot that correlates relative mobility (log R_f) to gel concentration (%T) was used to obtain a slope (K_r) for each protein standard (Figure 3.4B). These data were used to construct a Ferguson plot of –log K_r versus log molecular mass. By using these standard curves, the molecular mass of truncated P4H was estimated to be 140 kDa, which is in gratifying agreement with the results of analytical gel filtration chromatography. Heretofore, we refer to this truncated P4H as P4Hα(I)235–534/β.

Catalysis by Truncated P4H. The prolyl hydroxylase activity of P4Hα(I)235–534/β was determined by measuring the release of $[^{14}C]CO_2$ from $[1\text{-}^{14}C]$α-ketoglutarate during the hydroxylation of $(Pro–Pro–Gly)_{10}$ (10–1000 μM) and dansyl-Gly–Phe–Pro–Gly-OEt (10–1000 μM) peptide substrates (Table 2). The truncated enzyme has weak binding constants for both peptide substrates, and the K_M for $(Pro–Pro–Gly)_{10}$ (0.96 mM) is greater than the K_M for the tetrapeptide (0.23 mM). The rate of hydroxylation is an order of magnitude slower for the tetrapeptide than for $(Pro–Pro–Gly)_{10}$, although saturating conditions could not be reached due to the limited solubility of the substrate peptides.

---

[1] If the (actual mass)/(predicted mass) ratio is similar for the full-length and truncated P4H, then the actual mass of the truncated P4H is (228 kDa/380 kDa) × 140 kDa = 84 kDa, which is most consistent with the truncated P4H being an αβ dimer (87 kDa). This stoichiometry has been observed in *C. elegans* P4H (Myllyharju & Kivirikko, 1997).

Production of Full-Length P4H. We tested two characteristics of the P4H expression vector as being responsible for the production of a truncated α subunit. First, we suspected that the translation initiation region for P4Hα(I) in the pBK1.PDI1.P4H5 vector had significant secondary structure that could hinder ribosome binding. To address this possibility, we constructed a plasmid (pBK1.PDI1.P4H6) with the same translation initiation region, taken directly from plasmid pET22b(+), for both PDI and P4Hα(I). The 32-kDa P4Hα(I) was still expressed from pBK1.PDI1.P4H6 (data not shown) suggesting that ribosome binding was not responsible for the truncation.

Second, we noted that P4Hα(I) contains nine methionine residues that could serve as alternative start codons. Translation initiation at Met235 would produce a *C*-terminal α subunit of 32.2 kDa, in agreement with the truncated α subunit observed by SDS–PAGE. To address this possibility, we used site-directed mutagenesis to replace Met235 with a leucine residue (which has a similar hydrophobicity and volume), yielding plasmid pBK1.PDI1.P4H7.

By making this final mutation to the bicistronic expression vector, we were able to produce full-length α subunit in Origami™ B(DE3) cells. SDS–PAGE analysis of crude whole cell extract revealed the absence of an induced 32-kDa band (Figure 3.5A). The full-length α and β subunits were of similar molecular mass (55.3 kDa and 59.8 kDa, respectively), and hence could not be distinguished easily by Coomassie staining. Immunoblot analysis did, however, reveal the induction of both subunits in whole cell extracts (Figure 3.5B and 3.5C). Apparently, Met235 provides an alternative start codon for the translation of P4Hα(I).

Production of P4H from plasmid pBK1.PDI1.P4H7 was also attempted in Rosetta-gami™ 2(DE3) cells. These cells combine an oxidizing cytosol with *t*RNA's that are rare in *E. coli*. Induction with 500 μM IPTG at 23 °C yielded no P4H according to SDS–PAGE analysis (data not shown).

P4H expressed from plasmid pBK1.PDI1.P4H7 bound tightly to a poly(proline)-affinity column (Figure 3.5A). Purified enzyme was eluted with a 5 mg/mL solution of poly(proline), and then applied to an anion-exchange column, which also served to separate poly(proline) from the enzyme. P4H eluted as a single peak at 250 mM NaCl. A final gel-filtration step yielded protein that was >90% pure as estimated by SDS–PAGE, which represents a 148-fold purification over the ammonium sulfate fraction (Table 3.1). Approximately 2 mg of P4H tetramer was obtained per L of growth medium (41% yield).

Catalysis by Full-Length P4H. Kinetic analysis of the full-length P4H enzyme produced in E. coli was performed by measuring the release of $[^{14}C]CO_2$ from $[1\text{-}^{14}C]\alpha$-ketoglutarate during the hydroxylation of $(Pro\text{–}Pro\text{–}Gly)_{10}$ (2–200 μM) and dansyl-Gly–Phe–Pro–Gly-OEt (10–500 μM) peptide substrates at 30 °C (Table 3.2). $(Pro\text{–}Pro\text{–}Gly)_{10}$ binds more than 80-fold more tightly to full-length enzyme than to the truncated enzyme, whereas the tetrapeptide substrate binds 3-fold more tightly to the full-length than to the truncated enzyme. The k_{cat} for hydroxylation of both substrates is approximately 10 min^{-1}. This rate is approximately 60-fold less than that reported for the hydroxylation of $(Pro\text{–}Pro\text{–}Gly)_{10}$ at 37 °C by P4H tetramer produced in a baculovirus system (Vuori *et al.*, 1992a). The reported K_M value for this substrate is 18 μM (Vuori *et al.*, 1992b), which agrees well with that measured herein for P4H produced in *E. coli*.

HPLC-Based Assay for Enzymatic Activity. An HPLC-based assay was developed to measure the *in vitro* enzymatic activity of P4H enzyme. This method was based on a previously published substrate (Tandon *et al.*, 1998; Wu *et al.*, 1999) that had been monitored for hydroxylation by thin-layer chromatography. Here, the fluorescent tetrapeptide substrate (dansyl-Gly–Phe–Pro–Gly-OEt) was resolved from its hydroxylated product (dansyl-Gly–Phe–Hyp–Gly-OEt) during reversed-phase HPLC, requiring only a 15-min isocratic elution with acetonitrile/water (1:1) (Figure 3.6). The identity of each peak was confirmed independently by LC/MS analysis (data not shown).

The validity of our hydroxylation assay was confirmed by comparing it to the traditional CO_2-release assay (Table 3.3). The K_M value for the turnover of the tetrapeptide substrate by purified full-length human P4H tetramer is 94 μM by the hydroxylation assay, which is within error of the K_M value obtained with the CO_2-release assay (87 μM). The k_{cat} value of 19 min^{-1} obtained by HPLC analysis is twofold greater than that obtained with the CO_2-release assay (9.2 min^{-1}).

3.5 Discussion

Recombinant human P4H tetramer was successfully expressed and assembled from a bicistronic expression vector in Origami™ B(DE3) cells. This *E. coli* expression strain contains mutations in the thioredoxin reductase (*trxB*) and glutathione reductase (*gor*) genes, which enhance the ability of the cells to form protein disulfide bonds in the *E. coli* cytosol (Prinz *et al.*, 1997; Bessette *et al.*, 1999a). Previous experiments have revealed that P4H subunits can be dissociated by reduction with dithiothreitol *in vivo* and *in vitro* (Berg & Prockop, 1973a; Nietfeld *et al.*, 1981). Studies of P4H assembly in a cell-free system, using translated α-subunit mRNA and endogenous PDI, showed that a properly assembled α subunit contains an intramolecular disulfide bond (John & Bulleid, 1994). Site-directed mutagenesis of the five conserved α-subunit cysteine residues has shown that Cys276 and Cys293 form an intramolecular disulfide bond that is essential for tetramer assembly (John & Bulleid, 1994) and that Cys486 and Cys511 are required for assembly of active tetramers, possibly through an intramolecular disulfide bond (John & Bulleid, 1994; Lamberg *et al.*, 1995). *In vivo,* P4H dissociated by dithiothreitol can be reassociated oxidatively through interactions with the cellular chaperone BiP (John & Bulleid, 1996). *In vitro,* however, it has not been possible to reassociate P4H subunits (Nietfeld *et al.*, 1981; Koivu & Myllylä, 1986). In the reducing environment of the *E. coli* cytosol, it is likely that the intramolecular disulfide bond(s) of the α subunit cannot form, precluding it from assembling with PDI into the $\alpha_2\beta_2$ tetramer (as observed in BL21(DE3) cells). Allowing P4Hα(I) to form disulfide bond(s) in the relatively oxidizing cytosol of Origami™ B(DE3) cells enables tetramer assembly in the absence of any other eukaryotic proteins.

Surprisingly, induction of the initial bicistronic expression vector (pBK1.PDI1.P4H5) in Origami™ B(DE3) cells resulted in the production of a truncated α subunit. This subunit is synthesized from an alternative start codon (Met235) of P4Hα(I), and folds in the relatively oxidizing cytosol of the Origami™ B(DE3) cells but not the reducing cytosol of BL21(DE3) cells. Furthermore, under these conditions, this plasmid is unable to direct the production of full-length α subunit. Met235 is conserved among P4Hα(I) subunits from human, mouse, rat, and chicken (Bassuk *et al.*, 1989; Helaakoski *et al.*, 1989; Hopkinson *et al.*, 1994; Helaakoski *et al.*, 1995), but is replaced by a leucine residue in the α(II) and α(III) isoforms of these organisms

(Helaakoski *et al.*, 1995; Annunen *et al.*, 1997; Kukkola *et al.*, 2003). Whether translation at Met235 plays any natural role is unknown.

The 32-kDa α subunit that was synthesized by Origami™ B(DE3) cells contains only the *C*-terminal residues 235–534, but retains the ability to catalyze prolyl hydroxylation (Table 3.2). A number of site-directed mutagenesis studies have shown that the catalytic residues of P4H are positioned at the *C*-terminus of the α subunit. His412, Asp414, and His483 are the iron-binding ligands (Lamberg *et al.*, 1995; Myllyharju & Kivirikko, 1997), while Lys493 and His501 bind α-ketoglutarate (Myllyharju & Kivirikko, 1997). All of these residues are present in P4Hα(I)235–534/β. The binding of collagen substrates, on the other hand, involves residues that are more *N*-terminal than Met235. Limited proteolysis studies have revealed that the peptide-binding site is distinct from the catalytic domain and is composed of residues 140–240 (Myllyharju & Kivirikko, 1999). Indeed, a Phe144–Ser244 polypeptide has a K_d value for (Pro–Pro–Gly)$_{10}$ that is nearly identical to the K_M value for the turnover of (Pro–Pro–Gly)$_{10}$ by the P4H tetramer (Hieta *et al.*, 2003). These residues are predominantly missing from the truncated α subunit produced from plasmid pBK1.PDI1.P4H5. Studies with model peptides have suggested that the binding of collagen substrates to P4H involves the β-turn of a Pro–Gly sequence preceded by a polyproline type-II helix (Atreya & Ananthanarayanan, 1991). In addition to binding one proline residue in the active site for catalysis, the enzyme binds the extended strand in additional subsites that likely exist in the more *N*-terminal portion of the α subunit. The inability of the P4Hα(I)235–534 subunit to bind to a poly(proline)-affinity column supports this location for its substrate binding subsites.

Kinetic studies reveal that P4Hα(I)235–534/β binds substrates more weakly than does the full-length P4H tetramer (Table 3.2). In addition, the truncated α subunit binds (Pro–Pro–Gly)$_{10}$ even more weakly than dansyl-Gly–Phe–Pro–Gly–OEt, in opposition to the trend observed for full-length P4H (Table 3.2). Typically, P4H binds longer substrates more tightly than shorter substrates (Kivirikko *et al.*, 1972), presumably due to additional favorable interactions with subsites of the α subunit.

By removing the alternative start codon, full-length human P4H tetramer was produced in Origami™ B(DE3) cells and purified to >90% homogeneity. The recombinant enzyme has kinetic properties that are comparable to P4H produced in other heterologous systems. The K_M value measured here for (Pro–Pro–Gly)$_{10}$ is 11 μM, which is similar to that measured for P4H purified from a baculovirus system (18 μM) (Vuori *et al.*, 1992b). The k_{cat} for hydroxylation of (Pro–Pro–Gly)$_{10}$ is lower for P4H purified in this study (13 min^{-1}) than that reported for P4H produced in baculovirus (12.5–13 s^{-1}) (Vuori *et al.*, 1992b). A difference in reaction temperature (30 °C for *E. coli*-produced P4H and 37 °C for baculovirus-produced P4H) could contribute to this difference.

The HPLC-based hydroxylation assay described herein has advantages over the traditional CO_2-release assay for P4H activity. CO_2(g) release necessarily provides an indirect measurement of hydroxylation, which can be problematic. In the absence of sufficient substrate, P4H mediates the uncoupled decarboxylation of α-ketoglutarate, which does not produce Hyp and inactivates the enzyme by an as yet unidentified mechanism (Myllylä *et al.*, 1984; Wu *et al.*, 2000). Even under conditions of saturating substrate, this uncoupled reaction occurs to a small degree. For this reason, ascorbate is included in all activity assays to rescue enzyme trapped in the Fe(III) state. The HPLC-based assay provides a direct measurement of hydroxylation, thereby avoiding complications arising from the uncoupled reaction. Moreover, the HPLC-based assay avoids the use of radioactivity and the complexities of trapping a gaseous product. Finally, the CO_2-release assay method requires comparatively large amounts of enzyme, as the measurement of $[^{14}C]CO_2$ release requires that each time point derive from a separate reaction mixture. In contrast, the HPLC-based assay allows for the removal of aliquots from a single reaction mixture.

The HPLC-based assay has additional advantages over less common assays that monitor hydroxylation directly by using radiolabeled procollagen substrates (Juva & Prockop, 1966; Kivirikko & Myllylä, 1982). These assays require the tedious separation of ^{14}C- or ^{3}H-labeled hydroxyproline from radiolabeled proline prior to scintillation counting. In contrast, the HPLC-based assay allows for the direct analysis of the reaction mixture in an automated manner.

To assess the accuracy of the new assay, kinetic constants obtained for hydroxylation of a tetrapeptide substrate by the HPLC-based assay were compared to those from the more traditional CO_2-release assay. The results obtained provide validation for its continued use. The K_M values obtained by the two assays are in gratifying agreement. The difference in k_{cat} values could be explained (at least in part) by incomplete trapping of $[^{14}C]CO_2$ during the CO_2-release assay, providing yet another benefit to the assay developed here.

Collagen is a common biomaterial (Werkmeister & Ramshaw, 1992; Ramshaw *et al.*, 1996). Bovine collagen, which is most often used in this context, can engender allergic and immunological side effects in humans, as well as other health risks. Collagen has been produced previously in the yeast *Pichia pastoris* by the coexpression of cDNAs for P4H and procollagen (Vuorela *et al.*, 1997). Likewise, the expression system described herein could be used to produce natural human collagen in *E. coli*.

Table 3.1 Purification of full-length tetrameric prolyl 4-hydroxylase[a]

Purification step	Total protein (mg)	Activity[b] (10^3 units)	Specific activity (units/mg)	Yield (%)	Purification (fold)
30–65% ammonium sulfate fraction	591	9.2	15.5	100	1
anion-exchange chromatography	2.0	3.7	1880	40	121
gel-filtration chromatography	1.5	3.8	2300	41	148

[a] Data are for a 1.0-L culture of Origami™ B(DE3) cells grown in TB medium. Data for the poly(proline)-affinity chromatography are not shown because poly(proline) that leaches from the resin is a potent inhibitor of the enzyme.

[b] One unit of prolyl 4-hydroxylase is the amount required to hydroxylate 1 μM dansyl-Gly–Phe–Pro–Gly-OEt to form dansyl-Gly–Phe–Hyp–Gly-OEt per min.

Table 3.2 Kinetic parameters for prolyl hydroxylation by P4Hα(I)235–534/β and full-length P4H[a]

	P4Hα(I)235–534/β[b]		full-length P4H[c]	
parameter	$(PPG)_{10}$	dansyl-GFPG-OEt	$(PPG)_{10}$	dansyl-GFPG-OEt
K_M (mM)	0.96 ± 0.43	0.23 ± 0.06	0.011 ± 0.002	0.078 ± 0.032
V_{max} (μM min^{-1})	4.5 ± 1.2	0.48 ± 0.05	4.7 ± 0.2	3.3 ± 0.4
k_{cat} (min^{-1})	4.5 ± 1.2	0.48 ± 0.05	13 ± 1	9.2 ± 1.2
k_{cat}/K_M ($M^{-1}s^{-1}$)	78 ± 46	35 ± 14	$(2.0 \pm 0.6) \times 10^4$	$(1.8 \pm 0.6) \times 10^3$

[a] Data were obtained by measuring $[^{14}C]CO_2$ released during the decarboxylation of $[1\text{-}^{14}C]\alpha$-ketoglutarate. Reactions were performed in Tris (50 mM, pH 7.8) at 30 °C for 15 min. Reaction components were as described in the text. dansyl-Gly–Phe–Pro–Gly-OEt (10–1000 μM) or $(Pro–Pro–Gly)_{10}$ (2–1000 μM) were used as substrates.
[b] Purified P4Hα(I)235–534/β (90 μg/mL) was used in the reactions. Calculations for k_{cat} are for an αβ dimer.
[c] Purified full-length P4H (82 μg/mL) was used in the reactions. Calculations for k_{cat} are for an $\alpha_2\beta_2$ tetramer.

Table 3.3 Comparison of HPLC-based assay to [^{14}C]CO_2-release assay[a]

Assay method	K_M (μM)	V_{max} (μM·min^{-1})	k_{cat} (min^{-1})	k_{cat}/K_M (M^{-1}s^{-1})
[^{14}C]CO_2 release[b]	78 ± 32	3.3 ± 0.4	9.2 ± 1.2	(1.8 ± 0.6) × 10^3
HPLC[c]	94 ± 19	6.8 ± 0.5	19 ± 1.5	(3.3 ± 1.3) × 10^3

[a] Reactions were performed in 50 mM Tris–HCl buffer (pH 7.8) at 30 °C for 15 min. Reaction components were as described in the text. Dansyl-Gly–Phe–Pro–Gly-OEt substrate (10–1000 μM) and full-length P4H (82 μg/mL) purified from *E. coli* were used in both assays.

[b] [1-^{14}C]α-ketoglutarate (100 μM, 185000 dpm) was added to initiate the reaction.

[c] Peptide substrate was added to initiate the reaction.

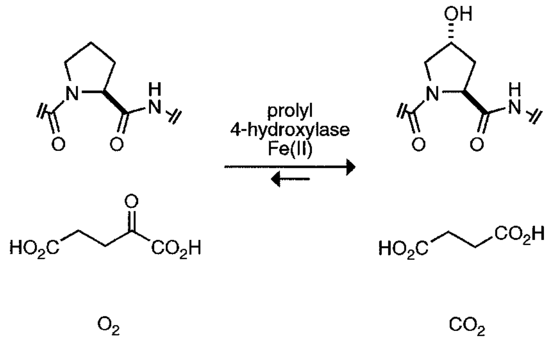

prolyl
4-hydroxylase
Fe(II)
O
HO2C
CO2H
O2
OH
N
H
CO2

Figure 3.1 Reaction catalyzed by prolyl 4-hydroxylase.

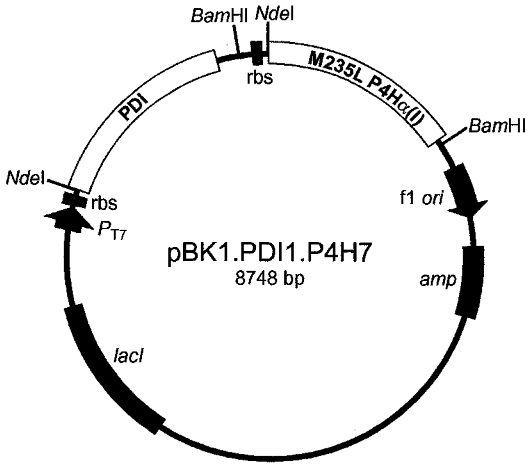

BamHI
NdeI
rbs
M235L P4Hα(I)
BamHI
PDI
NdeI
rbs
P_{T7}
f1 ori
pBK1.PDI1.P4H7
8748 bp
amp
lacI

Figure 3.2 Map of plasmid pBK1.PDI1.P4H7, which is a bicistronic expression vector in which cDNAs encoding human PDI and the M235L variant of human P4Hα(I) have been inserted into pET22b(+). Both cDNAs are transcribed from a single T7 promoter. Each resulting transcript contains an rbs for the initiation of translation.

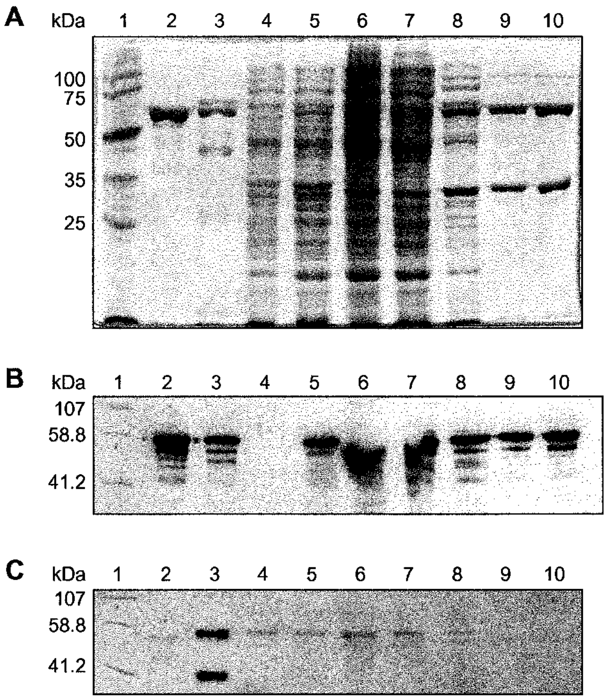

A
kDa
1 2 3 4 5 6 7 8 9 10
100
75
50
35
25
B
kDa
1 2 3 4 5 6 7 8 9 10
107
58.8
41.2
C
kDa
1 2 3 4 5 6 7 8 9 10
107
58.8
41.2

Figure 3.3 10% SDS–PAGE analysis of the production and purification of P4H from pBK1.PDI1.P4H5 in Origami™ B(DE3) cells, which results in a truncated α subunit. Lane 1, molecular-mass standards; lane 2, human PDI standard; lane 3, human P4H standard; lane 4, uninduced whole cells; lane 5, induced whole cells; lane 6, crude cell lysate; lane 7, 30–65% ammonium sulfate fraction; lane 8, pooled fractions from cation-exchange chromatography; lane 9, pooled fractions from anion-exchange chromatography; lane 10, pooled fractions from gel-filtration chromatography. **A**. Analysis by Coomassie staining. **B**. Analysis by immunoblotting using an anti-human PDI polyclonal antibody. **C**. Analysis by immunoblotting using an anti-human α(I) subunit monoclonal antibody, which does not recognize the truncated α subunit.

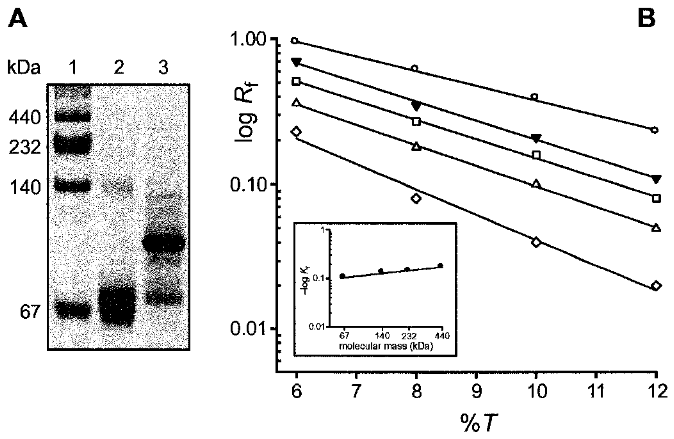


Figure 3.4 Native PAGE analysis of purified truncated P4H. **A**. Representative Coomassie-stained gel containing 7.5% (w/v) poly(acrylamide). Lane 1, molecular-mass standards; lane 2, human PDI standard; lane 3, truncated P4H, expressed from pBK1.PDI1.P4H5 in Origami™ B(DE3) cells and purified. **B**. Log plot of polyacrylamide concentration (%T) versus relative mobility (R_f). BSA (circles, 67 kDa, $K_r = -0.101$), lactate dehydrogenase (squares, 140 kDa, $K_r = -0.132$), catalase (open triangles, 232 kDa, $K_r = -0.141$), and ferritin (diamonds, 440 kDa, $K_r = -0.174$) were used to construct a log plot (inset) of molecular mass versus $-K_r$. Log plot of truncated P4H (filled triangles) gives $K_r = -0.132$, which corresponds to a molecular mass of 140 kDa.

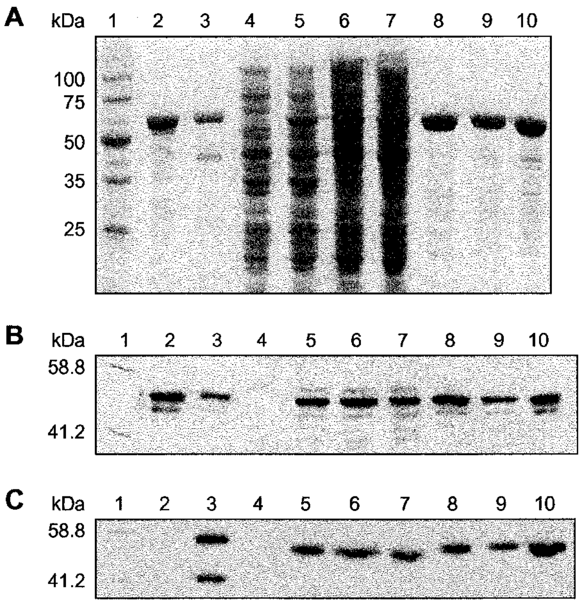

A
kDa
1 2 3 4 5 6 7 8 9 10
100
75
50
35
25
B
kDa
1 2 3 4 5 6 7 8 9 10
58.8
41.2
C
kDa
1 2 3 4 5 6 7 8 9 10
58.8
41.2

Figure 3.5 10% SDS–PAGE analysis of the production and purification of recombinant human P4H from pBK1.PDI1.P4H7 in Origami™ B(DE3) cells. Lane 1, molecular-mass standards; lane 2, human PDI standard; lane 3, human P4H standard; lane 4, uninduced whole cells; lane 5, induced whole cells; lane 6, crude cell lysate; lane 7, 30–65% ammonium sulfate fraction; lane 8, pooled fractions from poly(proline)-affinity chromatography; lane 9, pooled fractions from anion-exchange chromatography; lane 10, pooled fractions from gel-filtration chromatography. **A**. Analysis by Coomassie staining. **B**. Analysis by immunoblotting using an anti-human PDI polyclonal antibody. **C**. Analysis by immunoblotting using an anti-human α(I) subunit monoclonal antibody.

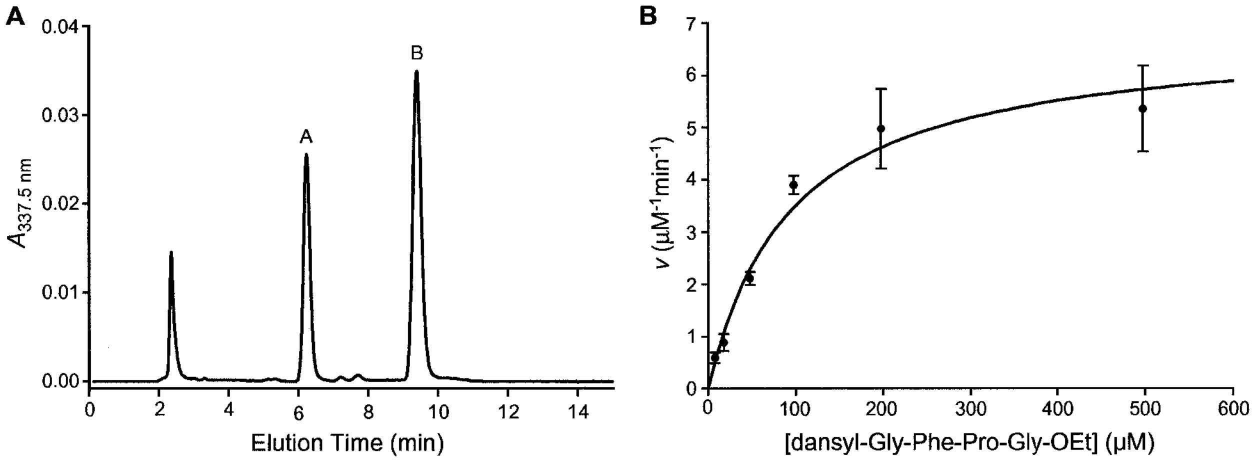


Figure 3.6 HPLC-based assay for P4H activity. **A**. HPLC trace of a time point in a representative assay of Dns-Gly–Phe–Pro–Gly-OEt hydroxylation. Enzymatic reactions were injected onto an analytical reversed-phase C18 column, and peptides were eluted isocratically with acetonitrile/water (1:1). Peak A is the Hyp-containing tetrapeptide product. Peak B is the Pro-containing tetrapeptide substrate. **B**. Catalysis of Dns-Gly–Phe–Pro–Gly-OEt hydroxylation by full-length human P4H. Reactions mixtures (100 μL) were 50 mM Tris–HCl buffer (pH 7.8) containing BSA (1 mg/mL), catalase (0.1 mg/mL), DTT (100 μM), ascorbate (2 mM), $FeSO_4$ (50 μM), P4H (0.36 μM), α-ketoglutarate (100 μM), and tetrapeptide substrate (10–500 μM). Reactions were performed at 30 °C for 15 min and quenched by boiling for 30 s. Aliqouts (20 μL) were analyzed as in panel **A**, with Peak A being quantified by integration of $A_{337.5\text{ nm}}$. Individual points are the average (± SE) of three independent hydroxylation reactions. To obtain kinetic parameters, data were fitted to the Michaelis–Menten equation: $v = V_{\max}[S]/(K_M + [S])$.

Chapter Four

Contribution of Tertiary Amides to the Conformational Stability of Collagen Triple Helices

This chapter was published as:

Kersteen, E. A., and Raines, R. T. (2001) Contribution of tertiary amides to the conformational stability of collagen triple helices. *Biopolymers* **59** 24–28.

4.1 Abstract

The collagen triple helix is composed of three polypeptide strands, each with a sequence of repeating (Xaa-Yaa-Gly) triplets. In these triplets, Xaa and Yaa are often tertiary amides: L-proline (Pro) and 4(*R*)-hydroxy-L-proline (Hyp). To determine the contribution of tertiary amides to triple-helical stability, Pro and Hyp were replaced in synthetic collagen mimics with a nonnatural acyclic tertiary amide: *N*-methyl-L-alanine (meAla). Replacing a Pro or Hyp residue with meAla decreases triple-helical stability. Ramachandran analysis indicates that meAla residues prefer to adopt ϕ and ψ angles that are dissimilar from those of the Pro and Hyp residues in the collagen triple helix. Replacement with meAla decreases triple-helical stability more than does replacement with Ala. All of the peptide bonds in triple-helical collagen are in the *trans* conformation. Although an Ala residue greatly prefers the *trans* conformation, a meAla residue exists as a nearly equimolar mixture of *trans* and *cis* conformers. These findings indicate that the favorable contribution of Pro and Hyp to the conformational stability of collagen triple helices arises from factors other than their being tertiary amides.

4.2 Introduction

Collagen is the most abundant protein in animals. The tertiary structure of collagen is a right-handed helix composed of three chains, each resembling a polyproline type-II helix. The primary sequence of each polypeptide chain is comprised of repeating Xaa-Yaa-Gly triplets, where Gly is glycine and Xaa and Yaa are often L-proline (Pro) and 4(*R*)-hydroxy-L-proline (Hyp), respectively. Together, Pro and Hyp comprise nearly $^1/_4$ of the residues in type I collagen, which is the most common type in humans (Fietzek & Kuhn, 1975). The abundance of these cyclic tertiary amides correlates with the conformational stability of the collagen triple helix (Josse & Harrington, 1964; Shah *et al.*, 1996).

Pro-Hyp-Gly is the most common triplet in type I collagen (Fietzek & Kuhn, 1975). Yet, Pro is not an especially common residue in other proteins, comprising 5.1% of all residues in known proteins (McCaldon & Argos, 1988). In some proline-rich proteins, Nature seems to have chosen Pro simply because it is the only naturally occurring residue that forms a tertiary amide. For example, Src homology 3 (SH3) and WW domains retain high-specificity recognition for proline-rich ligands when proline in these ligands is replaced with the nonnatural amino acid *N*-methylglycine (sarcosine; Sar) (Nguyen *et al.*, 1998). In other words, proline is recognized by SH3 and WW domains because it lacks a main-chain NH. Inspired by this finding, we decided to probe the role of the prevalent tertiary amides in the conformational stability of collagen.

The effect of tertiary amides on collagen stability is unclear from previous studies. Peptides in which Sar has been substituted at the Yaa position, poly(Gly-Pro-Sar), do not form stable triple helices (Goodman *et al.*, 1996; Goodman *et al.*, 1998). But like Gly, Sar has much conformational flexibility (Armand *et al.*, 1997). In contrast, *N*-isobutylglycine (Nleu) has been used in place of Pro and Hyp without sacrificing stability. Both $(\text{Gly-Nleu-Pro})_9$ and (Gly-Pro-

Nleu)$_9$ form more stable triple helices than does (Gly-Pro-Pro)$_{10}$, and Nleu is more stabilizing in the Xaa position than in the Yaa position (Goodman *et al.*, 1998; Kwak *et al.*, 1999). Hydrophobic interactions between the Nleu side chain and the Pro ring contribute to this stability (Melacini *et al.*, 1996; Feng *et al.*, 1997).

Here, we use synthetic collagen mimics to reveal the contribution of tertiary amides to the conformational stability of collagen triple helices. Three residues, Pro, *N*-methyl-L-alanine (meAla), and L-alanine (Ala) were placed in the Xaa and Yaa positions of the central triplet of a synthetic (Pro-Hyp-Gly)$_3$(Xaa-Yaa-Gly)(Pro-Hyp-Gly)$_3$ peptide. Each peptide forms a triple helix. Thermal denaturation experiments reveal that meAla destabilizes the triple helix more than does Ala, and that meAla is more destabilizing in the Xaa position than in the Yaa position (Figure 4.1).

4.3 Materials and Methods

Peptide Synthesis. Triple-helical peptides were prepared with the sequence (Pro-Hyp-Gly)$_3$(Xaa-Yaa-Gly)(Pro-Hyp-Gly)$_3$ in which the (Xaa-Yaa-Gly) triplet was (meAla-Hyp-Gly), (Pro-meAla-Gly), (Ala-Hyp-Gly), (Pro-Ala-Gly), and (Pro-Pro-Gly). Peptides were prepared by solid-phase synthesis on a Perkin-Elmer/Applied Biosystems Model 432A synthesizer using standard Fmoc chemistry. Fmoc-protected amino acids (Nova-Biochem, San Diego, CA) were coupled to Fmoc-glycine Wang resin (Advanced ChemTech, Louisville, KY). The peptides were deprotected and cleaved from the resin with trifluoroacetic acid (TFA) containing triisopropylsilane (TIS; 2.5% v/v) and water (2.5% v/v), and then precipitated with diethyl ether.

Peptides were purified by reversed-phase high performance liquid chromatography (RP-HPLC) using a Waters system (486 detector, 510 pumps) and a Pharmacia C-18 semipreparatory column. Peptides were eluted with a gradient of aqueous acetonitrile (5–40% v/v) containing TFA (0.1% v/v) at a flow rate of 3 mL/min. Purified peptides were obtained by collecting the major peak with absorbance at 215 nm, which occurred between 17 and 19% v/v acetonitrile. HPLC-purified peptides were analyzed by electrospray mass spectrometry using a PE Sciex API 365 triple quadrupole with an ionspray source, or by matrix-assisted laser desorption/ionization (MALDI) mass spectrometry using a Bruker REFLEX II (Billerica, MA) equipped with a 337 nm laser, reflectron, and delayed extraction.

Triple Helix Formation. Triple helices were formed by incubating peptides in 50 mM acetic acid (0.2 mM final concentration) for 24 h at 4 °C. Concentrations of peptides were determined by measuring absorbance at 214 nm ($\varepsilon = 4.48 \times 10^4$ $M^{-1}cm^{-1}$). Collagen has a characteristic circular dichroism (CD) spectrum that contains a peak at 225 nm. Triple helix formation was assessed at 5 °C by CD spectrometry on an Aviv 62A DS or an Aviv 202 SF instrument, both of which are equipped with an automated temperature controller.

Thermal Denaturation. Values for T_m for each triple helix were determined in triplicate by thermal denaturation experiments monitored by CD spectroscopy on an Aviv 62A DS or an Aviv 202 SF instrument, both equipped with an automated temperature controller. Ellipticity at 225 nm was monitored as the temperature was increased from 5–50 °C in 3°-C increments with a 3- or 5-min equilibration time at each temperature. As the temperature was increased, the ellipticity decreased. Data were fitted to a two-state model for unfolding to determine the value of T_m, which is the temperature at the midpoint of the thermal transition.

4.4 Results and Discussion

The γ carbon of proline is particularly fateful for the conformational stability of triple-helical collagen. Electron-withdrawing substituents on the γ carbon can increase stability. For example, a (Pro-Hyp-Gly)$_{10}$ triple helix has a T_m of 58 °C, whereas a (Pro-Pro-Gly)$_{10}$ triple helix has a T_m of only 24 °C (Sakakibara *et al.*, 1973). Replacing the hydroxyl group of Hyp with fluorine, the most electronegative atom, increases further the conformational stability of triple-helical collagen (Holmgren *et al.*, 1998; Holmgren *et al.*, 1999). An alternative approach to revealing the role of C_γ in collagen stability is to remove it. L-Azetidine-2-carboxylic acid (Aze) is a Pro analog that lacks a carbon, as it has a four-membered ring. Both poly(Gly-Pro-Aze) and poly(Gly-Aze-Pro) form less stable triple helices than does poly(Gly-Pro-Pro) (Zagari *et al.*, 1990). Here, we take a more subtle approach to removing C_γ while retaining a tertiary amide—replacing Pro with meAla.

Five triple-helical collagen-like peptides were synthesized by solid-phase methods, and the integrity of each was confirmed by mass spectrometry (data not shown). There was no evidence of diketopiperizine formation, based on HPLC analysis and mass spectrometry. (Xaa-Yaa-Gly) triplets of the sequence (meAla-Hyp-Gly), (Pro-meAla-Gly), (Ala-Hyp-Gly), (Pro-Ala-Gly), and (Pro-Pro-Gly) were introduced into the middle of a (Pro-Hyp-Gly)$_3$(Xaa-Yaa-Gly)(Pro-Hyp-Gly)$_3$ peptide. Circular dichroism spectroscopy indicated that each peptide formed a triple helix at low temperatures (data not shown). At a concentration of 0.2 mM in 50 mM acetic acid, the CD spectrum for each peptide contained the positive peak at 225 nm that is characteristic of the collagen triple helix.

With increasing temperature, the ellipticity of a collagen triple helix at 225 nm decreases as the concentration of the triple helix decreases. The resulting T_m value provide an indication of the conformational stability of a triple helix. The values of T_m for the five (Pro-Hyp-Gly)$_3$(Xaa-Yaa-Gly)(Pro-Hyp-Gly)$_3$ triple helices are affected significantly by the central triplet, increasing in the order: (meAla-Hyp-Gly) < (Pro-meAla-Gly) < (Ala-Hyp-Gly), (Pro-Ala-Gly) < (Pro-Pro-Gly) (Table 4.1; Figure 4.1). All five of these triple helices have a T_m significantly lower than that of (Pro-Hyp-Gly)$_7$. Substitution of meAla is highly destabilizing in both the Xaa and Yaa positions, and is more destabilizing in the Xaa position. Ala substitution is also destabilizing, in agreement with previous studies (Shah *et al.*, 1996), but less so than meAla substitution. There is no significant difference in T_m between Ala substitution in the Xaa versus Yaa position. Finally, a (Pro-Pro-Gly) triplet destabilizes the (Pro-Hyp-Gly)$_7$ triple helix, but less so than a triplet containing either Ala or meAla.

Why is meAla so destabilizing to a collagen triple helix? Instability could result from the unfavorable conformation of meAla. Structural data have shown that the average values for the dihedral angles of Pro in the (Pro-Hyp-Gly)$_{10}$ triple helix are $\phi = -72°$ and $\psi = 161°$, and of Hyp in a (Pro-Hyp-Gly)$_{10}$ triple helix are $\phi = -58°$ and $\psi = 152°$(Nagarajan *et al.*, 1999). Empirically-derived conformational energy maps for *N*-methylated analogues of three tripeptide hormones show that *N*-methyl substitution greatly restricts the conformational freedom of the ϕ and ψ dihedral angles (Manavalan & Momany, 1980). Ramachandran plots indicate that the most favorable dihedral angles for *N*-methylated residues in the tripeptides are approximately $\phi = -130°$ and $\psi = 70°$. These preferred angles are far from those of Pro and Hyp in a collagen triple helix. In addition, the average ϕ and ψ values for Pro and Hyp in collagen lie completely outside the broad low-energy region (bounded by the 10 kcal/mol contour) for *N*-methylated residues. In comparison to meAla, Ala has a broad low-energy region in tripeptides (Manavalan & Momany, 1980). Although the most favorable dihedral angles for Ala are approximately $\phi = -80°$ and $\psi = 80°$, the average ϕ and ψ angles for Pro and Hyp in $(\text{Pro-Hyp-Gly})_{10}$ lie well within the energy contour bounded by 10 kcal/mol.

Substitution of meAla decreases the stability of the triple helix more than does substitution of Ala. In part, the difference may be due to the tendency of meAla to adopt a *cis* peptide bond conformation. All of the peptide bonds in the collagen triple helix are in the *trans* conformation, and *cis* to *trans* isomerization is the rate-limiting step in the folding of a triple helix (Buevich *et al.*, 2000). Although Ala exists almost exclusively in the *trans* conformation, meAla exists predominantly in the *cis* conformation, at least in organic solvents (Vitoux *et al.*, 1986). These studies provide an explanation for our finding that meAla is more destabilizing to the collagen triple helix than is Ala. Interestingly, a Pro-Sar dipeptide has a lower percentage of *cis* conformers (15–35%) than does a Pro-meAla dipeptide (Vitoux *et al.*, 1986). *N*-Substituted glycine residues (such as Nleu) might therefore be less destabilizing than meAla due to a lower tendency to adopt the *cis* conformation.

The hydrogen bonding pattern in triple-helical collagen is strictly conserved between the amide N-H of glycine and the amide oxygen of the Xaa residue (Bella & Berman, 1996; Nagarajan *et al.*, 1999). It is reasonable that meAla substitution in the Xaa position causes conformational changes that weaken the interchain hydrogen bonds more than does substitution in the Yaa position. This reasoning can explain meAla substitution being significantly more destabilizing in the Xaa position ($T_m = 17.5$ °C) than in the Yaa position ($T_m = 21.7$ °C). Weaker hydrogen bonds caused by meAla substitution in the Xaa position would translate into a less stable triple helix.

4.5 Conclusion

In collagen, unlike in the ligands of SH3 and WW recognition domains, a tertiary amide is not enough. The conformational stability of collagen relies on more than the mere presence of tertiary amides. Instead, the conformational restrictions imposed by the pyrrolidine ring of Pro and Hyp are critical for the structural integrity of the collagen triple helix.

Table 4.1 Values of T_m for synthetic (Pro-Hyp-Gly)$_3$(Xaa-Yaa-Gly)(Pro-Hyp-Gly)$_3$ triple helices

Xaa-Yaa-Gly	T_m (°C)[a]
Pro-Hyp-Gly[b]	36 ± 2
Pro-Pro-Gly	30.5 ± 2.2
Ala-Hyp-Gly	26.1 ± 0.7
Pro-Ala-Gly	25.0 ± 1.0
Pro-meAla-Gly	21.7 ± 0.8
meAla-Hyp-Gly	17.5 ± 1.6

[a] Values of T_m were determined by circular dichroism spectroscopy for peptides (0.2 mM) in 50 mM acetic acid, and are the average (±SE) of at least 3 determinations.

[b] From (Bretscher *et al.*, 2001).

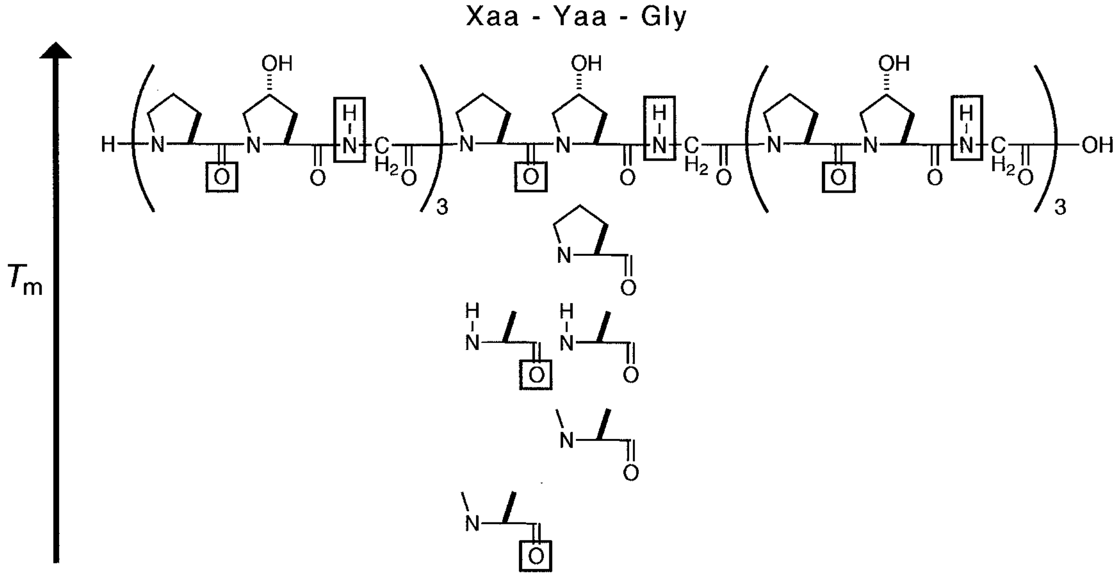


Figure 4.1 Relative conformational stability of (Pro-Hyp-Gly)$_3$(Xaa-Yaa-Gly)(Pro-Hyp-Gly)$_3$ triple helices in which the central triplets are (Pro-Hyp-Gly), (Pro-Pro-Gly), (Ala-Hyp-Gly), (Pro-Ala-Gly), (Pro-meAla-Gly), or (meAla-Hyp-Gly). Boxed atoms participate in interstrand hydrogen bonds in a collagen triple helix.

Appendix I

Phas-I as a Substrate for Human Prolyl 4-Hydroxylase

AI.1 Introduction

Insulin and other growth factors stimulate protein synthesis by the phosphorylation of a number of translation factors. One of these factors is the Phas-I (phosphorylated heat- and acid-stable) protein (also known as 4E-BP1), which regulates cap-dependent translation initiation (Sachs & Varani, 2000). Phas-I regulates the activity of eukaryotic initiation factor (eIF) 4E, the protein that binds to the 5′ mRNA cap structure, through changes in its phosphorylation state (von der Haar *et al.*, 2004). Unphosphorylated Phas-I binds tightly to eIF4E and disrupts the binding of eIF4E to other proteins in the multisubunit cap-binding protein complex. Phosphorylation of Phas-I, which increases upon insulin signaling (Hu *et al.*, 1994), greatly reduces its ability to bind to eIF4E. The weakening of the Phas-I/eIF4E interaction, in turn, allows formation of the cap-binding complex and initiation of translation.

The phosphorylation sites of Phas-I have been identified as residues Thr36, Thr45, Ser64, Thr69, Ser82, and Ser111 ((Fadden *et al.*, 1997); Figure AI.1). Thr36 and Thr45 are phosphorylated by a FRAP/mTOR (FKBP12-rapamycin associated protein/mammalian target of rapamycin)-associated pathway. Although this phosphorylation event can occur when Phas-I is in a complex with eIF4E, phosphorylation does not cause the dissociation of that complex (Gingras *et al.*, 1999). The phosphorylation of Thr36 and Thr45 is not dependent on external stimuli, but may be required for the phosphorylation of additional Phas-I residues in an insulin-sensitive manner (Gingras *et al.*, 1999). Another early phosphorylation event is carried out at Ser111 by a novel fat-cell kinase that is stimulated more than 10-fold by insulin (Heesom *et al.*, 1998). Phosphorylation of Ser64 is responsible for the largest decrease in affinity of Phas-I for eIF4E (Karim *et al.*, 2001).

The sites phosphorylated by FRAP/mTOR (Thr36 and Thr45) are followed directly by the sequence Pro-Gly-Gly (Figure AI.1). These amino acids resemble the Pro-Pro-Gly recognition sequence of the enzyme collagen prolyl 4-hydroxylase (P4H). P4H catalyzes the post-translational hydroxylation of proline residues in collagen (Myllyharju, 2003), a key step in the stabilization of the collagen triple helix (Chopra & Ananthanarayanan, 1982). We have attempted to hydroxylate Phas-I using human P4H to investigate whether P4H could play a role in regulating the phosphorylation of Phas-I.

AI.2 Materials and Methods

Human P4H was produced in *Escherichia coli* from a bicistronic expression vector and purified according to published procedures (Kersteen *et al.*, 2004). Recombinant rat Phas-I was obtained from Calbiochem (San Diego, CA) and dialyzed against 50 mM Tris–HCl buffer (pH 7.6) before being tested as a substrate in a standard hydroxylation assay. The assay was performed in 100 μL of 50 mM Tris–HCl buffer (pH 7.6) containing bovine serum albumin (1 mg/mL), catalase (100 μg/mL), DTT (100 μM), ascorbate (2 mM), $FeSO_4$ (50 μM), P4H (0.34 μM), and α-ketoglutarate (100 μM). The reaction was initiated by the addition of Phas-I (to 20 μM), and the reaction mixture was incubated at 30 °C for 6 h. Every 2 h, fresh P4H was added, such that the final P4H concentration was 0.91 μM.

The reaction mixture was analyzed for Phas-I hydroxylation by in-gel trypsin digestion followed by MALDI–MS, performed at the University of Wisconsin Biotechnology Center. Samples of dialyzed but unreacted Phas-I (2.5 μg) and the Phas-I hydroxylation reaction (2.1 μg) were subjected to 10% SDS–PAGE and then visualized by Coomassie staining (Figure AI.2).

Protein bands were excised from the gel, destained, and digested with trypsin. MALDI–MS was used to detect peptides isolated from the digest.

AI.3 Results

The unreacted Phas-I sample and the "hydroxylated" sample had nearly identical peptide products after trypsin digestion (Table AI.1). MALDI–MS revealed two peptides in the unreacted sample that were not obtained with the hydroxylated sample ($[M+H]^+$ of 1469.0 and 1783.3). The peptide of mass 1469.0 corresponds to residues 105–117 of rat Phas-1, which does not contain a Pro residue. The peptide of mass 1783.3 could not be identified. There are no peptides in the hydroxylated sample that do not appear in the unreacted sample, suggesting that none of the Pro residues in Phas-I had been hydroxylated by human P4H. There are, however, other prolyl hydroxylases that could catalyze the hydroxylation of proline residues in Phas-I in humans (Myllyharju, 2003).

Table AI.1 Results from MALDI–MS of trypsin-digested rat Phas-I

Unreacted[a] $[M+H]^+$	Hydroxylated[b] $[M+H]^+$	Phas-I residues	Predicted $[M+H]^+$	Amino acid sequence
807.6	807.7	51-56	806.5	IIYDRK
855.5	855.7	57-62	854.5	FLMECR (carbamidomethyl C)
869.6	871.7	57-62	870.5	FLMECR (carbamidomethyl C and oxidized M)
1469.0		105-117	1468.0	RAGGEESQFEMDI
1651.1	1651.4	n.d.		
1769.3	1769.6	n.d.		
1783.3		n.d.		
1797.4	1797.6	n.d.		
1811.4	1811.7	n.d.		
3020.1	3021.0	20-50	3019.1	VALGDGVQLPPGDYSTTPGGTLFSTTPGGTR
3501.0	3501.8	n.d.		
3542.3	3543.6	73-105	3541.5	DLPTIPGVTSPTSDEPPMQASQSHLHSSPEDKR
3558.4	3559.4	73-105	3557.5	DLPT....DKR (oxidized M)
3924.5	3925.4	n.d.		
3966.8	3967.2	69-105	3964.8	TPPKDLPTIPGVTSPTSDEPPMQASQSHLHSSPEDKR
3981.6	3982.9	69-105	3980.8	TPPK....DKR (oxidized M)

n.d. indicates mass values that were not matched to any Phas-I sequence.

[a] $[M+H]^+$ values obtained from trypsin digestion of dialyzed, unreacted rat Phas-I.

[b] $[M+H]^+$ values obtained from trypsin digestion of rat Phas-I reacted with human P4H.

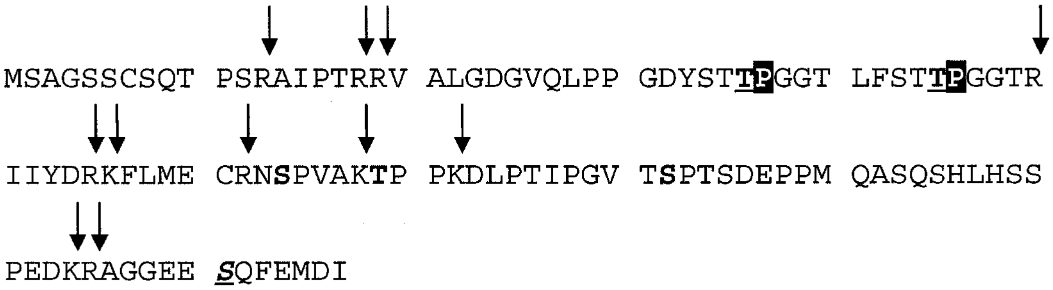

MSAGSSCSQT PSRAIPTRRV ALGDGVQLPP GDYSTTPGGT LFSTTPGGTR
IIYDRKFLME CRNSPVAKTP PKDLPTIPGV TSPTSDEPPM QASQSHLHSS
PEDKRAGGEE SQFEMDI

Figure AI.1 Sequence of rat Phas-I. FRAP/mTOR kinase sites are underlined. Ser111 (shown in italics) is phosphorylated by a novel fat-cell kinase and casein kinase II. Additional sites of phosphorylation are in bold. Potential P4H recognition residues are shown in white letters with a black background. Trypsin recognition sites are indicated with arrows.

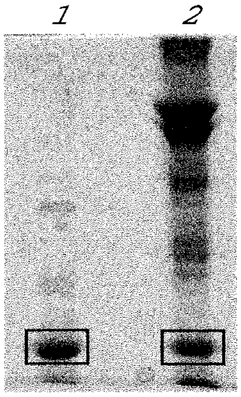


Figure AI.2 10% SDS–PAGE analysis of recombinant rat Phas-I. Lane 1, unreacted Phas-I; lane 2, reaction of Phas-I with P4H. Bands surrounded by a box were excised from the gel for trypsin digestion.

References

Akaji, K. (1997). Disulfide bond formation using silyl chloride-sulfoxide system. *Rev. Heteroatom Chem.* **16**, 85-100.

Akaji, K., Tatsumi, T., Yoshida, M., Kimura, T., Fujiwara, Y. & Kiso, Y. (1992). Disulfide bond formation using the silyl chloride-sulfoxide system for the synthesis of a cystine peptide. *J. Am. Chem. Soc.* **114**, 4137-4143.

Ala-Kokko, L., Baldwin, C. T., Moskowitz, R. W. & Prockop, D. J. (1990). Single base mutation in the type II procollagen gene (COL2A1) as a cause of primary osteoarthritis associated with a mild chondrodysplasia. *Proc. Natl. Acad. Sci. USA* **87**, 6565-6568.

Alanen, H. I., Williamson, R. A., Howard, M. J., Lappi, A.-K., Jäntti, H. P., Rautio, S. M., Kellokumpu, S. & Ruddock, L. W. (2003). Functional characterization of ERp18, a new endoplasmic reticulum-located thioredoxin superfamily member. *J. Biol. Chem.* **278**, 28912-28920.

Andersen, D. C. & Krummen, L. (2002). Recombinant protein expression for therapeutic applications. *Curr. Opin. Biotechnol.* **13**, 117-123.

Anelli, T., Alessio, M., Mezghrani, A., Simmen, T., Talamo, F., Bachi, A. & Sitia, R. (2002). ERp44, a novel endoplasmic reticulum folding assistant of the thioredoxin family. *EMBO J.* **21**, 835-844.

Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science* **181**, 223-230.

Annunen, P., Helaakoski, T., Myllyharju, J., Veijola, J., Pihlajaniemi, T. & 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.

Armand, P., Kirshenbaum, K., Falicov, A., Dunbrack, R. L., Jr., Dill, K. A., Zuckermann, R. N. & Cohen, F. E. (1997). Chiral N-substituted glycine can form stable helical conformations. *Fold. Des.* **2**, 369-375.

Atreya, P. L. & 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.

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, R. L. (1995). The nature of protein folding pathways: The classical versus the new view. *J. Biomolec. NMR* **5**, 103-109.

Baneyx, F. (1999). Recombinant protein expression in *Escherichia coli*. *Curr. Opin. Biotechnol.* **10**, 411-421.

Barbouche, R., Miquelis, R., Jones, I. M. & Fenouillet, E. (2003). Protein-disulfide isomerase-mediated reduction of two disulfide bonds of HIV envelope glycoprotein 120 occurs post-CXCR4 binding and is required for fusion. *J. Biol. Chem.* **278**, 3131-3136.

Bassuk, J. A., Kao, W. W.-Y., Herzer, P., Kedersha, N. L., Seyer, J., DeMartino, J. A., Daugherty, B. L., Mark III, G. E. & Berg, R. A. (1989). Prolyl 4-hydroxylase: Molecular cloning and the primary structure of the α subunit from chicken embryo. *Proc. Natl. Acad. Sci. USA* **86**, 7382-7386.

Bella, J. & Berman, H. M. (1996). Crystallographic evidence for Cα-H...O=C hydrogen bonds in a collagen triple helix. *J. Mol. Biol.* **264**, 734-742.

Bella, J., Eaton, M., Brodsky, B. & Berman, H. M. (1994). Crystal and molecular structure of a collagen-like peptide at 1.9 Å resolution. *Science* **266**, 75-81.

Berg, R. A. & Prockop, D. J. (1973a). Affinity column purification of protocollagen proline hydroxylase from chick embryos and further characterization of the enzyme. *J. Biol. Chem.* **218**, 1175-1182.

Berg, R. A. & Prockop, D. J. (1973b). Purification of [^{14}C]protocollagen and its hydroxylation by prolyl-hydroxylase. *Biochemistry* **12**, 3395-3401.

Berg, R. A. & 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. *Biochemistry Biophys. Res. Commun.* **52**, 115-120.

Berisio, R., Vitagliano, L., Mazzarella, L. & 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. & Zagari, A. (2002). Crystal structure of the collagen triple helix model $[(\text{Pro-Pro-Gly})_{10}]_3$. *Protein Sci.* **11**, 262-270.

Bessette, P. H., Åslund, F., Beckwith, J. & Georgiou, G. (1999a). Efficient folding of proteins with multiple disulfide bonds in the *Escherichia coli* cytoplasm. *Proc. Natl. Acad. Sci. USA* **96**, 13703-13708.

Bessette, P. H., Cotto, J. J., Gilbert, H. F. & Georgiou, G. (1999b). *In vivo* and *in vitro* function of the *Escherichia coli* periplasmic cysteine oxidoreductase DsbG. *J. Biol. Chem.* **274**, 7784-7792.

Borowski, T., Bassan, A. & Siegbahn, P. E. M. (2004). Mechanism of dioxygen activation in 2-oxoglutarate-dependent enzymes: A hybrid DFT study. *Chem. Eur. J.* **10**, 1031-1041.

Braig, K., Otwinowski, Z., Hegde, R., Boisvert, D. C., Joachimiak, A., Horwich, A. L. & Sigler, P. B. (1994). The crystal structure of the bacterial chaperonin GroEL at 2.8 Å. *Nature* **371**, 578-586.

Bretscher, L. E., Jenkins, C. L., Taylor, K. M., DeRider, M. L. & Raines, R. T. (2001). Conformational stability of collagen relies on a stereoelectronic effect. *J. Am. Chem. Soc.* **123**, 777-778.

Brodsky, J. (1998). Translocation of proteins across the endoplasmic reticulum membrane. *Int. Rev. Cytol.* **178**, 277-328.

Buevich, A. V., Dai, Q.-H., Liu, X., Brodsky, B. & Baum, J. (2000). Site-specific NMR monitoring of cis-trans isomerization in the folding of the proline-rich collagen triple helix. *Biochemistry* **39**, 4299-4308.

Bukau, B., Deuerling, E., Pfund, C. & Craig, E. A. (2000). Getting newly synthesized proteins into shape. *Cell* **101**, 119-122.

Bulleid, N. J., Wilson, R. & 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. *Biochemistry J.* **317**, 195-202.

Burjanadze, T. V. (2000). New analysis of the phylogenetic change of collagen thermostability. *Biopolymers* **53**, 523-528.

Burns, J. A. & Whitesides, G. M. (1990). Predicting the stability of cyclic disulfides by molecular modeling: "Effective concentrations" in thiol–disulfide interchange and the design of strongly reducing thiols. *J. Am. Chem. Soc.* **112**, 6296-6303.

Byers, P. H. (1993). Osteogenesis imperfecta. In *Connective tissue and its heritable disorders: molecular, genetic, and medical aspects* (Royce, P. M. & Steinmann, B., eds.). Wiley Liss, New York.

Byers, P. H. (2001). Folding defects in fibrillar collagens. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **356**, 151-158.

Cabibbo, A., Pagani, M., Fabbri, M., Rocchi, M., Farmery, M., Bulleid, N. & Sitia, R. (2000). ERO1-L, a human protein that favors disulfide bond formation in the endoplasmic reticulum. *J. Biol. Chem.* **275**, 4827-4833.

Cabrele, C., Fiori, S., Pegorano, S. & Moroder, L. (2002). Redox-active cyclic bis(cysteinyl) peptides as catalysts for in vitro oxidative protein folding. *Chem. Biol.* **9**, 731-740.

Cai, H., Wang, C. C. & Tsou, C. L. (1994). Chaperone-like activity of protein disulfide isomerase in the refolding of a protein with no disulfide bonds. *J. Biol. Chem.* **269**, 24550-24552.

Cardinale, G. J., Rhoads, R. E. & Udenfriend, S. (1971). Simultaneous incorporation of ^{18}O into succinate and hydroxyproline catalyzed by collagen proline hydroxylase. *Biochemistry Biophys. Res. Commun.* **43**, 537-543.

Cheung, P. Y. & Churchich, J. E. (1999). Recognition of protein substrates by protein-disulfide isomerase. *J. Biol. Chem.* **274**, 32757-32761.

Chivers, P. T., Laboissière, M. C. A. & Raines, R. T. (1996). The CXXC motif: Imperatives for the formation of native disulfide bonds in the cell. *EMBO J.* **16**, 2659-2667.

Chivers, P. T., Laboissière, M. C. A. & Raines, R. T. (1998). Protein disulfide isomerase: Cellular enzymology of the CXXC motif. In *Prolyl Hydroxylase, Protein Disulfide Isomerase, and Other Structurally-Related Proteins* (Guzman, N. A., ed.), pp. 487-505. Marcel Dekker, New York.

Chivers, P. T., Prehoda, K. E. & Raines, R. T. (1997a). The CXXC motif: A rheostat in the active site. *Biochemistry* **36**, 4061-4066.

Chivers, P. T., Prehoda, K. E., Volkman, B., Kim, B.-M., Markley, J. L. & Raines, R. T. (1997b). Microscopic pK_a values of *Escherichia coli* thioredoxin. *Biochemistry* **36**, 14985-14991.

Chivers, P. T. & Raines, R. T. (1997). General acid/base catalysis in the active site of *Escherichia coli* thioredoxin. *Biochemistry* **36**, 15810-15816.

Chopra, R. K. & Ananthanarayanan, V. S. (1982). Conformational implications of enzymatic proline hydroxylation in collagen. *Proc. Natl. Acad. Sci. USA* **79**, 7180-7184.

Christiansen, C., St. Hilaire, P. M. & Winther, J. R. (2004). Fluorometric polyethyleneglycol-peptide hybrid substrates for quantitative assay of protein disulfide isomerase. *Anal. Biochemistry* **333**, 148-155.

Chu, L. & Robinson, D. K. (2001). Industrial choices for protein production by large-scale cell culture. *Curr. Opin. Biotechnol.* **12**, 180-187.

Cleland, W. W. (1964). Dithiothreitol, a new protective reagent for SH groups. *Biochemistry* **3**, 480-482.

Clifton, I. J., Hsueh, L.-C., Baldwin, J. E., Harlos, K. & Schofield, C. J. (2001). Structure of proline 3-hydroxylase. *Eur. J. Biochemistry* **268**, 6625-6636.

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

Counts, D. F., Cardinale, G. J. & Udenfriend, S. (1978). Prolyl hydroxylase half reaction: peptidyl prolyl-independent decarboxylation of α-ketoglutarate. *Proc. Natl. Acad. Sci. USA* **75**, 2145-2149.

Cuthbertson, A. & Indrevoll, B. (2000). A method for the one-pot regioselective formation of the two disulfide bonds of α-conotoxin SI. *Tetrahedron Lett.* **41**, 3661-3663.

D'Andrea, L. D. & Regan, L. (2003). TPR proteins: the versatile helix. *Trends Biochem. Sci.* **28**, 655-662.

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

Darby, N. J. & Creighton, T. E. (1995a). Characterization of the active site cysteine residues of the thioredoxin-like domains of protein disulfide isomerase. *Biochemistry* **34**, 16770-16780.

Darby, N. J. & Creighton, T. E. (1995b). Functional properties of the individual thioredoxin-like domains of protein disulfide isomerase. *Biochemistry* **34**, 11725-11735.

Darby, N. J., Freedman, R. B. & Creighton, T. E. (1994). Dissecting the mechanism of protein disulfide isomerase: Catalysis of disulfide bond formation in a model peptide. *Biochemistry* **33**, 7937-7947.

Darby, N. J., Penka, E. & Vincentelli, R. (1998). The multi-domain structure of protein disulfide isomerase is essential for high catalytic efficiency. *J. Mol. Biol.* **276**, 239-247.

Darby, N. J., van Straaten, M., Penka, E., Vincentelli, R. & Kemmink, J. (1999). Identifying and characterizing a second structural domain of protein disulfide isomerase. *FEBS Lett.* **448**, 167-172.

de Jong, L., Albracht, S. P. J. & 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.

DeCollo, T. V. & Lees, W. J. (2001). Effects of aromatic thiols on thiol–disulfide interchange reactions that occur during protein folding. *J. Org. Chem.* **66**, 4244-4249.

Dill, K. A. & Chan, H. S. (1997). From Levinthal to pathways to funnels. *Nat. Struct. Biol.* **4**, 10-19.

Dobson, C. M. (2003). Protein folding and misfolding. *Nature* **426**, 884-890.

Dorner, A. J., Wasley, L. C., Raney, P., Haugejorden, S., Green, M. & Kaufman, R. J. (1990). The stress response in Chinese hamster ovary cells. Regulation of ERp72 and protein disulfide isomerase expression and secretion. *J. Biol. Chem.* **265**, 22029-22034.

Eis, P. S. & Lakowicz, J. R. (1993). Time-resolved energy transfer measurements of donor-acceptor distance distributions and intramolecular flexibility of a CCHH zinc finger peptide. *Biochemistry* **32**, 7981-7993.

Elkins, J. M., Hewitson, K. S., McNeill, L. A., Seibel, J. F., Schlemminger, I., Pugh, C. W., Ratcliffe, P. J. & Schofield, C. J. (2003). Structure of factor-inhibiting hypoxia-inducible factor (HIF) reveals mechanism of oxidative modification of HIF-1α. *J. Biol. Chem.* **278**, 1802-1806.

Elkins, J. M., Ryle, M. J., Clifton, I. J., Hotopp, J. C. D., Lloyd, J. S., Burzlaff, N. I., Baldwin, J. E., Hausinger, R. P. & Roach, P. L. (2002). X-ray crystal structure of *Escherichia coli* taurine/α-ketoglutarate dioxygenase complexed to ferrous iron and substrates. *Biochemistry* **41**, 5185-5192.

Ellgaard, L. & Helenius, A. (2003). Quality control in the endoplasmic reticulum. *Nat. Rev. Mol. Cell Biol.* **4**, 181-191.

Engel, J., Chen, H.-T., Prockop, D. J. & 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.

Engel, J. & Prockop, D. J. (1991). The zipper-like folding of collagen triple helices and the effects of mutations that disrupt the zipper. *Annu. Rev. Biophys. Biophys. Chem.* **20**, 137-152.

Ewalt, K. L., Hendrick, J. P., Houry, W. A. & Hartl, F. U. (1997). In vivo observation of polypeptide flux through the bacterial chaperonin system. *Cell* **90**, 491-500.

Eyre, D. R., Paz, M. A. & Gallop, P. M. (1984). Cross-linking in collagen and elastin. *Annu. Rev. Biochemistry* **53**, 717-748.

Fadden, P., Haystead, T. A. & Lawrence, J. C. J. (1997). Identification of phosphorylation sites in the traditional regulator, PHAS-I, that are controlled by insulin and rapamycin in rat adipocytes. *J. Biol. Chem.* **272**, 10240-10247.

Farquhar, R., Honey, H., Murant, S. J., Bossier, P., Schultz, L., Montgomery, D., Ellis, R. W., Freedman, R. B. & Tuite, M. F. (1991). Protein disulfide isomerase is essential for viability in *Saccharomyces cerevisiae*. *Gene* **108**, 81-89.

Feldman, D. E. & Frydman, J. (2000). Protein folding in vivo: the importance of molecular chaperones. *Curr. Opin. Struct. Biol.* **10**, 26-33.

Feng, Y., Melacini, G. & Goodman, M. (1997). Collagen-based structures containing the peptoid residue N-isobutylglycine (Nleu): synthesis and biophysical studies of Gly-Nleu-Pro sequences by circular dichroism and optical rotation. *Biochemistry* **36**, 8716-8724.

Ferbitz, L., Maier, T., Patzelt, H., Bukau, B., Deuerling, E. & Ban, N. (2004). Trigger factor in complex with the ribosome forms a molecular cradle for nascent proteins. *Nature* **431**, 590-596.

Fersht, A. R. (2000). Transition-state structure as a unifying basis in protein-folding mechanisms: Contact order, chain topology, stability, and the extended nucleus mechanism. *Proc. Natl. Acad. Sci. USA* **97**, 1525-1529.

Fessler, L. I., Morris, N. P. & Fessler, J. H. (1975). Procollagen: Biological scission of amino and carboxyl extension peptides. *Proc. Natl. Acad. Sci. USA* **72**, 4905-4909.

Fietzek, P. P. & Kuhn, K. (1975). Information contained in the amino acid sequence of the alpha1(I)-chain of collagen and its consequences upon the formation of the triple helix, of fibrils and crosslinks. *Mol. Cell. Biochemistry* **8**, 141-157.

Flynn, G. C., Chappell, T. G. & Rothman, J. E. (1989). Peptide binding and release by proteins implicated as catalysts of protein assembly. *Science* **245**, 385-390.

Flynn, G. C., Pohl, J., Flocco, M. T. & Rothman, J. E. (1991). Peptide-binding specificity of the molecular chaperone BiP. *Nature* **353**, 726-730.

Forman-Kay, J. D., Clore, G. M., Wingfield, P. T. & Gronenborn, A. M. (1991). High-resolution three-dimensional structure of reduced recombinant human thioredoxin in solution. *Biochemistry* **30**, 2685-2698.

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

Frand, A. R. & Kaiser, C. A. (1998). The *ERO1* gene of yeast is required for oxidation of protein dithiols in the endoplasmic reticulum. *Mol. Cell* **1**, 161-170.

Frand, A. R. & Kaiser, C. A. (1999). Ero1p oxidizes protein disulfide isomerase in a pathway for disulfide bond formation in the endoplasmic reticulum. *Mol. Cell* **4**, 469-477.

Frand, A. R. & Kaiser, C. A. (2000). Two pairs of conserved cysteines are required for the oxidative activity of Ero1p in protein disulfide bond formation in the endoplasmic reticulum. *Mol. Biol. Cell* **11**, 2833-2843.

Freedman, R. B., Hirst, T. R. & Tuite, M. F. (1994). Protein disulphide isomerase: Building bridges in protein folding. *Trends Biochemistry Sci.* **19**, 331-336.

Freedman, R. B., Klappa, P. & Ruddock, L. W. (2003). Model peptide substrates and ligands in analysis of action of mammalian protein disulfide-isomerase. *Methods Enzymol.* **348**, 342-354.

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

Gallina, A., Hanley, T. M., Mandel, R., Trahey, M., Broder, C. C., Viglianti, G. A. & Ryser, H. J.-P. (2002). Inhibitors of protein-disulfide isomerase prevent cleavage of disulfide bonds in receptor-bound glycoprotein 120 and prevent HIV-1 entry. *J. Biol. Chem.* **277**, 50579-50568.

Gilbert, H. F. (1990). Molecular and cellular aspects of thiol–disulfide exchange. *Adv. Enzymol.* **63**, 69-172.

Gilbert, H. F. (1997). Protein disulfide isomerase and assisted protein folding. *J. Biol. Chem.* **272**, 29399-29402.

Gilbert, H. F. (1998). Protein disulfide isomerase. *Methods Enzymol.* **290**, 26-50.

Gill, S. C. & von Hippel, P. H. (1989). Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochemistry* **182**, 319-326.

Gingras, A.-C., Gygi, S. P., Raught, B., Polakiewicz, R. D., Abraham, R. T., Hoekstra, M. F., Aebersold, R. & Sonenberg, N. (1999). Regulation of 4E-BP1 phosphorylation: a novel two-step mechanism. *Genes Dev.* **13**, 1422-1437.

Givol, D., Goldberger, R. F. & Anfinsen, C. B. (1964). Oxidation and disulfide interchange in the reactivation of reduced ribonuclease. *J. Biol. Chem.* **239**, PC3114-3116.

Goldberger, R. F., Epstein, C. J. & Anfinsen, C. B. (1963). Acceleration of reactivation of reduced bovine pancreatic ribonuclease by a microsomal system from rat liver. *J. Biol. Chem.* **238**, 628-635.

Goldstone, D., Haebel, P. W., Katzen, F., Bader, M. W., Bardwell, J. C., Beckwith, J. & Metcalf, P. (2001). DsbC activation by the N-terminal domain of DsbD. *Proc. Natl. Acad. Sci. USA* **98**, 9551-9556.

Goodman, M., Bhumralkar, M., Jefferson, E. A., Kwak, J. & Locardi, E. (1998). Collagen mimetics. *Biopolymers* **47**, 127-142.

Goodman, M., Melacini, G. & Feng, Y. (1996). Collagen-like triple helices incorporating peptoid residues. *J. Am. Chem. Soc.* **118**, 10928-10929.

Gough, J. D., Williams, J. R. H., Donofrio, A. E. & Lees, W. J. (2002). Folding disulfide-containing proteins faster with an aromatic thiol. *J. Am. Chem. Soc.* **124**, 3885-3892.

Goulding, C. W. & Perry, L. J. (2003). Protein production in *Escherichia coli* for structural studies by X-ray crystallography. *J. Struct. Biol.* **142**, 133-143.

Grauschopf, U., Winther, J. R., Korber, P., Zander, T., Dallinger, P. & Bardwell, J. C. A. (1995). Why is DsbA such an oxidizing disulfide catalyst? *Cell* **83**, 947-955.

Gross, E., Kastner, D. B., Kaiser, C. A. & Fass, D. (2004). Structure of Ero1p, source of disulfide bonds for oxidative protein folding in the cell. *Cell* **117**, 601-610.

Groves, J. T. & McClusky, G. A. (1976). Aliphatic hydroxylation via oxygen rebound. Oxygen transfer catalyzed by iron. *J. Am. Chem. Soc.* **98**, 859-861.

Gutsche, I., Essen, L.-O. & Baumeister, W. (1999). Group II chaperonins: new TRiC(k)s and turns of a protein folding machine. *J. Mol. Biol.* **293**, 295-312.

Guzman, N. A., Ed. (1998). *Prolyl Hydroxylase, Protein Disulfide Isomerase, and Other Structurally Related Proteins*. New York: Marcel Dekker.

Haebel, P. W., Goldstone, D., Katzen, F., Beckwith, J. & Metcalf, P. (2002). The disulfide bond isomerase DsbC is activated by an immunoglobulin-fold thiol oxidoreductase: Crystal structure of the DsbC–DsbDα complex. *EMBO J.* **21**, 4774-4784.

Hawkins, H. C. & Freedman, R. B. (1991). The reactivities and ionization properties of the active-site dithiol groups of mammalian protein disulphide-isomerase. *Biochem J.* **275**, 335-339.

Heesom, K. J., Avison, M. B., Diggle, T. A. & Denton, R. M. (1998). Insulin-stimulated kinase from rat fat cells that phosphorylates initiation factor 4E-binding protein 1 on the rapamycin-insensitive site (serine-111). *Biochem J.* **336**, 39-48.

Helaakoski, T., Annunen, P., Vuori, K., Macneil, I. A., Pihlajaniemi, T. & 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. & 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.

Hennecke, J., Sillen, A., Huber-Wunderlich, M., Engelborghs, Y. & Glockshuber, R. (1997). Quenching of tryptophan fluorescence by the active-site disulfide bridge in the DsbA protein from *Escherichia coli*. *Biochemistry* **36**, 6391-6400.

Heras, B., Edeling, M. A., Schirra, H. J., Raina, S. & Martin, J. L. (2004). Crystal structures of the DsbG disulfide isomerase reveal an unstable disulfide. *Proc. Natl. Acad. Sci. USA* **101**, 8876-8881.

Hieta, R., Kukkola, L., Permi, P., Pirilä, P., Kivirikko, K. I., Kipeläinen, I. & Myllyharju, J. (2003). The peptide-substrate-binding domain of human collagen prolyl 4-hydroxylases. Backbone assignments, secondary structure, and binding of proline-rich peptides. *J. Biol. Chem.* **278**, 34966-34974.

Higgins, S. J. & Hames, B. D. (1999). *Protein Expression: A Practical Approach*, Oxford University Press, Oxford, UK.

Ho, R. Y. N., Mehn, M. P., Hegg, E. L., Liu, A., Ryle, M. J., Hausinger, R. P. & Que, L. J. (2001). Resonance Raman studies of the iron(II)-α-keto acid chromophore in model and enzyme complexes. *J. Am. Chem. Soc.* **123**, 5022-5029.

Hodges, J. A. & Raines, R. T. (2003). Stereoelectronic effects on collagen stability: The dichotomy of 4-fluoroproline diastereomers. *J. Am. Chem. Soc.* **125**, 9263-9264.

Holmgren, S. K., Bretscher, L. E., Taylor, K. M. & Raines, R. T. (1999). A hyperstable collagen mimic. *Chem. Biol.* **6**, 63-70.

Holmgren, S. K., Taylor, K. M., Bretscher, L. E. & Raines, R. T. (1998). Code for collagen's stability deciphered. *Nature* **392**, 666-667.

Hopkinson, I., Smith, S. A., Donne, A., Gregory, H., Franklin, T. J., Grant, M. E. & Rosamond, J. (1994). The complete cDNA derived sequence of the rat prolyl 4-hydroxylase α subunit. *Gene* **149**, 391-392.

Hu, C., Pang, S., Kong, X., Vellaca, M. & Lawrence, J. C. J. (1994). Molecular cloning and tissue distribution of PHAS-I, an intracellular target for insulin and growth factors. *Proc. Natl. Acad. Sci. USA* **91**, 3730-3734.

Huber-Wunderlich, M. & Glockshuber, R. (1998). A single dipeptide sequence modulates the redox properties of a whole enzyme family. *Fold. Des.* **3**, 161-171.

Hunt, J. F., Weaver, A. J., Landry, S. J., Gierasch, L. & Deisenhofer, J. (1996). The crystal structure of the GroES co-chaperonin at 2.8Å resolution. *Nature* **379**, 37-42.

Hutton, J. J., Kaplan, A. & 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. & Udenfriend, S. (1966). Requirements for α-ketoglutarate, ferrous iron and ascorbate by collagen proline hydroxylase. *Biochemistry Biophys. Res. Commun.* **24**, 179-184.

Hwang, C., Sinskey, A. J. & Lodish, H. F. (1992). Oxidized redox state of glutathione in the endoplasmic reticulum. *Science* **257**, 1496-1502.

Inouye, K., Kobayashi, Y., Kyogoku, Y., Kishida, Y., Sakakibara, S. & Prockop, D. J. (1982). Synthesis and physical properties of (hydroxyproline–proline–glycine)$_{10}$. Hydroxyproline in the X-position decreases the melting temperature of the collagen triple helix. *Arch. Biochemistry Biophys.* **219**, 198-203.

Iwata-Reuyl, D., Basak, A. & Townsend, C. A. (1999). β-Secondary kinetic isotope effects in the clavaminate synthase-catalyzed oxidative cyclization of proclavaminic acid and in related azetidinone model reactions. *J. Am. Chem. Soc.* **121**, 11356-11368.

Jenkins, C. L. & Raines, R. T. (2002). Insights on the conformational stability of collagen. *Nat. Prod. Rep.* **19**, 49-59.

Joelson, T., Sjöberg, B. M. & Eklund, H. (1990). Modifications of the active center of T4 thioredoxin by site-directed mutagenesis. *J. Biol. Chem.* **265**, 3183-3188.

John, D. C. A. & Bulleid, N. J. (1994). Prolyl 4-hydroxylase: Defective assembly of the α-subunit mutants indicates that assembled α-subunits are intramolecularly disulfide bonded. *Biochemistry* **33**, 14018-14025.

John, D. C. A. & Bulleid, N. J. (1996). Intracellular dissociation and reassembly of prolyl 4-hydroxylase: The α-subunits associate with the immunoglobulin-heavy-chain binding protein (BiP) allowing reassembly with the β-subunit. *Biochemistry J.* **317**, 659-665.

John, D. C. A., Grant, M. E. & Bulleid, N. J. (1993). Cell-free synthesis and assembly of prolyl 4-hydroxylase: the role of the β-subunit (PDI) in preventing misfolding and aggregation of the α-subunit. *EMBO J.* **12**, 1587-1595.

Josse, J. & Harrington, W. F. (1964). Role of pyrrolidine residues in the structure and stabilization of collagen. *J. Mol. Biol.* **9**, 269-287.

Juva, K. & Prockop, D. J. (1966). Modified procedure for the assay of H3- or C14-labeled hydroxyproline. *Anal. Biochemistry* **15**, 77-83.

Kadler, K. E., Hojima, Y. & Prockop, D. J. (1987). Assembly of collagen fibrils *de novo* by cleavage of the type I pC-collagen with procollagen C-proteinase. Assay of critical concentration demonstrates that collagen self-assembly is a classical example of an entropy-driven process. *J. Biol. Chem.* **260**, 15696-15701.

Kadokura, H. & Beckwith, J. (2002). Four cysteines of the membrane protein DsbB act in concert to oxidize its substrate DsbA. *EMBO J.* **21**, 2354-2363.

Kamber, B., Hartmann, A., Eisler, K., Riniker, B., Rink, H., Sieber, P. & Rittel, W. (1980). The synthesis of cystine peptides by iodine oxidation of S-trityl-cysteine and S-acetamidomethyl-cysteine peptides. *Helvet. Chim. Acta* **63**, 899-915.

Karim, M. M., Hughes, J. M. X., Karwicker, J., Scheper, G. C., Proud, C. G. & McCarthy, J. E. G. (2001). A quantitative molecular model for modulation of mammalian translation by the eIF4E-binding protein 1. *J. Biol. Chem.* **276**, 20750-20757.

Kawano, K., Yoneya, T., Miyata, T., Yoshikawa, K., Tokunaga, F., Terada, Y. & Iwanaga, S. (1990). Antimicrobial peptide, tachyplesin I, isolated from hemocytes of the horseshoe crab (*Tachypleus tridentatus*). NMR determination of the β-sheet structure. *J. Biol. Chem.* **265**, 15365-15367.

Kemmink, J., Darby, N. J., Dijkstra, K., Nilges, M. & Creighton, T. E. (1996). Structure determination of the N-terminal thioredoxin-like domain of protein disulfide isomerase using multidimensional heteronuclear $^{13}C/^{15}N$ NMR spectroscopy. *Biochemistry* **35**, 7684-7691.

Kemmink, J., Darby, N. J., Dijkstra, K., Nilges, M. & Creighton, T. E. (1997). The folding catalyst protein disulfide isomerase is constructed of active and inactive thioredoxin modules. *Curr. Biol.* **7**, 239-245.

Kersteen, E. A., Higgin, J. J. & Raines, R. T. (2004). Production of human prolyl 4-hydroxylase in *Escherichia coli. Protein Expr. Purif.* **38**, 279-291.

Kersteen, E. A. & Raines, R. T. (2003). Catalysis of protein folding by protein disulfide isomerase and small-molecule mimics. *Antiox. Redox. Signal.* **5**, 413-424.

Kirsch, E., Krieg, T., Remberger, K., Fendel, H., Bruckner, P. & Muller, P. K. (1981). Disorder of collagen metabolism in a patient with osteogenesis imperfecta (lethal type): increased degree of hydroxylation of lysine in collagen types I and III. *Eur. J. Clin. Invest.* **11**, 39-47.

Kishore, R. & Balaram, P. (1985). Stabilization of γ-turn conformations in peptides by disulfide bridging. *Biopolymers* **24**, 2041-2043.

Kivirikko, K. I., Kishida, Y., Sakakibara, S. & Prockop, D. J. (1972). Hydroxylation of (X-Pro-Gly)$_n$ by protocollagen proline hydroxylase. Effect of chain length, helical conformation and amino acid sequence in the substrate. *Biochim. Biophys. Acta* **271**, 347-356.

Kivirikko, K. I. & Myllyharju, J. (1998). Prolyl 4-hydroxylases and their protein disulfide isomerase subunit. *Matrix Biol.* **16**, 357-368.

Kivirikko, K. I. & Myllyla, R. (1980). The hydroxylation of prolyl and lysyl residues. In *The enzymology of post-translational modification of proteins* (Freedman, R. B. & Hawkins, H. C., eds.), pp. 53-104. Academic Press, London.

Kivirikko, K. I. & Myllylä, R. (1982). Posttranslational enzymes in the biosynthesis of collagen: Intracellular enzymes. *Methods Enz.* **82**, 245-304.

Kivirikko, K. I. & Myllylä, R. (1987). Recent developments in posttranslational modification: Intracellular processing. *Methods Enz.* **144**, 96-114.

Kivirikko, K. I., Myllylä, R. & 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. & Pihlajaniemi, T. (1998). Collagen hydroxylases and the protein disulfide isomerase subunit of prolyl 4-hydroxylases. *Adv. Enzymol. Relat. Areas Mol. Biol.* **72**, 325-398.

Klappa, P., Hawkins, H. C. & Freedman, R. B. (1997). Interactions between protein disulphide isomerase and peptides. *Eur. J. Biochemistry* **248**, 37-42.

Klappa, P., Ruddock, L. W., Darby, N. J. & Freedman, R. B. (1998). The b′ domain provides the principal peptide-binding site of protein disulfide isomerase but all domains contribute to binding misfolded proteins. *EMBO J.* **17**, 927-935.

Koivu, J. & Myllylä, R. (1986). Protein disulfide-isomerase retains procollagen prolyl 4-hydroxylase structure in its native conformation. *Biochemistry* **25**, 5982-5986.

Koivu, J., Myllylä, R., Helaakoski, T., Pihlajaniemi, T., Tasanen, K. & Kivirikko, K. I. (1987a). Molecular cloning of the β-subunit of human prolyl-4-hydroxylase. This subunit and protein disulphide isomerase are products of the same gene. *EMBO J.* **6**, 643-649.

Koivu, J., Myllylä, R., Helaakoski, T., Pihlajaniemi, T., Tasanen, K. & Kivirikko, K. I. (1987b). A single polypeptide acts both as the β subunit of prolyl 4-hydroxylase and as a protein disulfide-isomerase. *J. Biol. Chem.* **262**, 6447-6449.

Koivunen, P., Pirneskoski, A., Karvonen, P., Ljung, J., Helaakoski, T., Notbohm, H. & Kivirikko, K. I. (1999). The acidic C-terminal domain of protein disulfide isomerase is not critical for the enzyme subunit function or for the chaperone or disulfide isomerase activities of the polypeptide. *EMBO J.* **18**, 65-74.

Koivunen, P., Salo, K. E. H., Myllyharju, J. & Ruddock, L. W. (2005). Three binding sites in protein disulfide isomerase co-operate in collagen prolyl 4-hydroxylase tetramer assembly. *J. Biol. Chem.* **280**, 5227-5235.

Konishi, Y., Ooi, T. & Scheraga, H. A. (1981). Regeneration of ribonuclease A from the reduced protein. Isolation and identification of intermediates, and equilibrium treatment. *Biochemistry* **20**, 3945-3955.

Kortemme, T. & Creighton, T. E. (1995). Ionisation of cysteine residues at the termini of model α-helical peptides. Relevance to unusual thiol pK_a values in proteins of the thioredoxin family. *J. Mol. Biol.* **253**, 799-812.

Kramer, R. Z., Bella, J., Mayville, P., Brodsky, B. & 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. & 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.

Krause, G., Lundström, J., Barea, J. L., Pueyo de la Cuesta, C. & Holmgren, A. (1991). Mimicking the active site of protein disulfide-isomerase by substitution of proline 34 in *Escherichia coli* thioredoxin. *J. Biol. Chem.* **266**, 9494-9500.

Kuivaniemi, H., Tromp, G. & Prockop, D. J. (1997). Mutations in fibrillar collagens (types I, II, III, and XI), fibril-associated collagen (type IX), and network-forming collagen (type X) cause a spectrum of diseases of bone, cartilage, and blood vessels. *Hum. Mutat.* **9**, 300-315.

Kukkola, L., Hieta, R., Kivirikko, K. I. & 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. & 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.

Kunkel, T. A. (1985). Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc. Natl. Acad. Sci. USA* **82**, 488-492.

Kwak, J., Jefferson, E. A., Bhumralkar, M. & Goodman, M. (1999). Triple helical stabilities of guest–host collagen mimetic structures. *Bioorg. Med. Chem.* **7**, 153-160.

Laboissière, M. C. A., Chivers, P. T. & Raines, R. T. (1995a). Production of rat protein disulfide isomerase in *Saccharomyces cerevisiae. Protein Express. Purif.* **6**, 700-706.

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

Laederach, A., Andreotti, A. H. & Fulton, D. B. (2002). Solution and micelle-bound structures of tachyplesin I and its active aromatic linear derivatives. *Biochemistry* **41**, 12359-12368.

Lakowicz, J. R., Gryczynski, I., Laczko, G., Wiczk, W. & Johnson, M. L. (2004). Distribution of distances between the tryptophan and the N-terminal residue of melittin in its complex with calmodulin, troponin C, and phospholipids. *Protein Sci.* **3**, 628-637.

Lakowicz, J. R., Gryczynski, I., Wiczk, W., Laczko, G., Prendergast, F. C. & Johnson, M. L. (1990). Conformational distributions of melittin in water/methanol mixtures from frequency-domain measurements of nonradiative energy transfer. *Biophys. Chem.* **36**, 99-115.

LaMantia, M., Miura, T., Tachikawa, H., Kaplan, H. W., Lennarz, W. J. & Mizunaga, T. (1991). Glycosylation site binding protein and protein disulfide isomerase are identical and essential for cell viability in yeast. *Proc. Natl. Acad. Sci. USA* **88**, 4453-4457.

Lamberg, A., Jauhiainen, M., Metso, J., Ehnholm, C., Shoulders, C., Scott, J., Pihlajaniemi, T. & Kivirikko, K. I. (1996). The role of protein disulphide isomerase in the microsomal triacylglycerol transfer protein does not reside in its isomerase activity. *Biochemistry J.* **315**, 533-536.

Lamberg, A., Pihlajaniemi, T. & 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.

Lang, K. & Scmid, F. X. (1988). Protein-disulphide isomerase and prolyl isomerase act differently and independently as catalysts of protein folding. *Nature* **331**, 453-455.

Lappi, A. K., Lensink, M. F., Alanen, H. I., Salo, K. E. H., Lobell, M., Juffer, A. H. & Ruddock, L. W. (2004). A conserved arginine plays a role in the catalytic cycle of the protein disulphide isomerases. *J. Mol. Biol.* **335**, 283-295.

Lee, C., Kim, S. J., Jeong, D. G., Lee, S. M. & 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, J. F. & Bulleid, N. J. (1994). The role of cysteine residues in the folding and association of the COOH-terminal propeptide of types I and III procollagen. *J. Biol. Chem.* **269**, 24354-24360.

Lees, J. F., Taseb, M. & Bulleid, N. J. (1997). Identification of the molecular recognition sequence which determines the type-specific assembly of procollagen. *EMBO J.* **16**, 908-916.

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

Lindberg, U., Schutt, C. E., Hellsten, E., Tjäder, A.-C. & Hult, T. (1988). The use of poly(L-proline)-sepharose in the isolation of profilin and profilactin complexes. *Biochim. Biophys. Acta* **967**, 391-400.

Liu, A., Ho, R. Y. N. & Que, L. J. (2001). Alternative reactivity of an α-ketoglutarate-dependent iron(II) oxygenase: enzyme self-hydroxylation. *J. Am. Chem. Soc.* **123**, 5126-5127.

Lundström, J. & Holmgren, A. (1993). Determination of the reduction-oxidation potential of the thioredoxin-like domains of protein disulfide-isomerase from the equilibrium with glutathione and thioredoxin. *Biochemistry* **32**, 6649-6655.

Lyles, M. M. & Gilbert, H. F. (1991). Catalysis of the oxidative folding of ribonuclease A by protein disulfide isomerase: Dependence of the rate on the composition of the redox buffer. *Biochemistry* **30**, 613-619.

Lyles, M. M. & Gilbert, H. F. (1994). Mutations in the thioredoxin sites of protein disulfide isomerase reveal functional nonequivalence of the N- and C-terminal domains. *J. Biol. Chem.* **269**, 30946-30952.

Lyman, S. K. & Schekman, R. (1997). Binding of secretory precursor polypeptides to a translocon subcomplex is regulated by BiP. *Cell* **88**, 85-96.

Majamaa, K., Günzler, V., Hanauske-Abel, H. M., Myllyla, R. & 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. & 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. Biochemistry* **138**, 239-245.

Maliwal, B. P., Lakowicz, J. R., Kupryszewski, G. & Rekowski, P. (1993). Fluorescence study of conformational flexibility of RNase S-peptide: Distance-distribution, end-to-end diffusion, and anisotropy decays. *Biochemistry* **32**, 12337-12345.

Manavalan, P. & Momany, F. A. (1980). Conformational energy studies on N-methylated analogs of thyrotropin releasing hormone, enkephalin, and luteinizing hormone-releasing hormone. *Biopolymers* **19**, 1943-1973.

Mao, J.-R. & Bristow, J. (2001). The Ehlers-Danlos syndrome: on beyond collagens. *J. Clin. Invest.* **107**, 1063-1069.

Marston, F. A. O. (1986). The purification of eukaryotic proteins synthesized in *Escherichia coli. Biochemistry J.* **240**, 1-12.

Matsuzaki, K., Yoneyama, S., Fujii, N., Miyajima, K., Yamada, K., Kirino, Y. & Anzai, K. (1997). Membrane permeabilization mechanisms of a cyclic antimicrobial peptide, tachyplesin I, and its linear analog. *Biochemistry* **36**, 9799-9806.

Matthias, L. J., Yam, P. T. W., Jiang, X.-M., Vandegraaff, N., Li, P., Poumbourios, P., Donoghue, N. & Hogg, P. J. (2002). Disulfide exchange in domain 2 of CD4 is required for entry of HIV-1. *Nature Immunol.* **3**, 727-710.

McCaldon, P. & Argos, P. (1988). Oligopeptide biases in protein sequences and their use in predicting protein coding regions in nucleotide sequences. *Proteins* **4**, 99-122.

McCarthy, A. A., Haebel, P. W., Törrönen, A., Rybin, V., Baker, E. N. & Metcalf, P. (2000). Crystal structure of the protein disulfide bond isomerase DsbC from *Escherichia coli. Nat. Struct. Biol.* **7**, 196-199.

Melacini, G., Feng, Y. & Goodman, M. (1996). Collagen-based structures containing the peptoid residue N-isobutylglycine (Nleu). 6. Conformational analysis of Gly-Pro-Nleu sequences by 1H NMR, CD, and molecular modeling. *J. Am. Chem. Soc.* **118**, 10725-10732.

Mezghrani, A., Fassio, A., Benham, A., Simmen, T., Braakman, I. & Sitia, R. (2001). Manipulation of oxidative protein folding and PDI redox state in mammalian cells. *EMBO J.* **20**, 6288-6296.

Molinari, M. & Helenius, A. (1999). Glycoproteins form mixed disulfides with oxidoreductases during folding in living cells. *Nature* **402**, 90-93.

Moroder, L., Besse, D., Musiol, H.-J., Rudolph-Böhner, S. & Siedler, F. (1996). Oxidative folding of cysteine-rich peptides vs regioselective cysteine pairing strategies. *Biopolymers (Peptide Sci.)* **40**, 207-234.

Müller, I., Kahnert, A., Pape, T., Sheldrick, G. M., Meyer-Klaucke, W., Dierks, T., Kertesz, M. & Usón, I. (2004). Crystal structure of the alkylsufatase AtsK: Insights into the catalytic mechanism of the Fe(II) α-ketoglutarate-dependent dioxygenase superfamily. *Biochemistry* **43**, 3075-3088.

Munro, S. & Pelham, H. R. B. (1987). A C-terminal signal prevents secretion of luminal ER proteins. *Cell* **48**, 899-907.

Myllyharju, J. (2003). Prolyl 4-hydroxylases, the key enzymes of collagen biosynthesis. *Matrix Biol.* **22**, 15-24.

Myllyharju, J. & 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. & Kivirikko, K. I. (1999). Identification of a novel proline-rich peptide-binding domain in prolyl 4-hydroxylase. *EMBO J.* **18**, 306-312.

Myllyharju, J. & Kivirikko, K. I. (2004). Collagens, modifying enzymes and their mutations in humans, flies and worms. *Trends Genet.* **20**, 33-43.

Myllylä, R., Kuutti-Savolainen, E.-R. & Kivirikko, K. I. (1978). The role of ascorbate in the prolyl hydroxylase reaction. *Biochemistry Biophys. Res. Commun.* **83**, 441-448.

Myllylä, R., Majamaa, K., Günzler, V., Hanauske-Abel, H. M. & 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., Schubotz, L. M., Weser, U. & Kivirikko, K. I. (1979). Involvement of superoxide in the prolyl and lysyl hydroxylase reactions. *Biochemistry Biophys. Res. Commun.* **89**, 98-102.

Myllyla, R., Tuderman, L. & Kivirikko, K. I. (1977). Mechanism of the prolyl hydroxylation reaction. 2. Kinetic analysis of the reaction sequence. *Eur. J. Biochemistry* **80**, 349-357.

Nagarajan, V., Kamitori, S. & Okuyama, K. (1998). Crystal structure analysis of collagen model peptide (Pro–Pro–Gly)$_{10}$. *J. Biochemistry* **124**, 1117-1123.

Nagarajan, V., Kamitori, S. & Okuyama, K. (1999). Structure analysis of a collagen-model peptide with a (Pro–Hyp–Gly) sequence repeat. *J. Biochemistry (Tokyo)* **125**, 310-318.

Nakamura, T., Furunaka, H., Miyata, T., Tokunaga, F., Muta, T., Iwanaga, S., Niwa, M., Takao, T. & Shimonishi, Y. (1988). Tachyplesin, a class of antimicrobial peptide from the hemocytes of the horseshoe crab (*Tachypleus tridentatus*). Isolation and chemical structure. *J. Biol. Chem.* **263**, 16709-16713.

Narayan, M., Welker, E., Wedemeyer, W. J. & Scheraga, H. A. (2000). Oxidative folding of proteins. *Acc. Chem. Res.* **33**, 805-812.

Nguyen, J. T., Turck, C. W., Cohen, F. E., Zuckermann, R. N. & Lim, W. A. (1998). Exploiting the basis of proline recognition by SH3 and WW domains: design of N-substituted inhibitors. *Science* **282**, 2088-2092.

Nietfeld, J. J. & 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. & Kemp, A. (1981). Dissociation and reassociation of prolyl 4-hydroxylase subunits after cross-linking of monomers. *Biochim. Biophys. Acta* **661**, 21-27.

Noiva, R., Freedman, R. B. & Lennarz, W. J. (1993). Peptide binding to protein disulfide isomerase occurs at a site distinct from the active sites. *J. Biol. Chem.* **268**, 19210-19217.

Nørgaard, P., Westphal, V., Tachibana, C., Alsoe, L., Holst, B. & Winther, J. R. (2001). Functional differences in yeast protein disulfide isomerases. *J. Cell. Biol.* **152**, 553-562.

Nørgaard, P. & Winther, J. R. (2001). Mutation of yeast Eug1p CXXS active sites to CXXC results in a dramatic increase in protein disulphide isomerase activity. *Biochemistry J.* **358**, 269-274.

Oishi, O., Yamashita, S., Nishimoto, E., Lee, S., Sugihara, G. & Ohno, M. (1997). Conformations and orientations of aromatic amino acid residues of tachyplesin I in phospholipid membranes. *Biochemistry* **36**, 4352-4359.

Oliver, J. D., Roderick, H. L., Llewellyn, D. H. & High, S. (1999). ERp57 functions as a subunit of specific complexes formed with the ER lectins calreticulin and calnexin. *Mol. Biol. Cell.* **10**, 2573-2582.

Østergaard, H., Henriksen, A., Hansen, F. G. & Winther, J. R. (2001). Shedding light on disulfide bond formation: Engineering a redox switch in green fluorescent protein. *EMBO J.* **20**, 5853-5862.

Ou, W.-J., Cameron, P. H., Thomas, D. Y. & Bergeron, J. J. M. (1993). Association of folding intermediates of glycoproteins with calnexin during protein maturation. *Nature* **364**, 771-776.

Pagani, M., Fabbri, M., Benedetti, C., Fassio, A., Pilati, S., Bulleid, N. J., Cabbibo, A. & Sitia, R. (2000). Endoplasmic reticulum oxidoreductin 1-Lβ (*ERO1-Lβ*), a human gene induced in the course of the unfolded protein response. *J. Biol. Chem.* **275**, 23685-23692.

Panasik, N., Jr., Eberhardt, E. S., Edison, A. S., Powell, D. R. & Raines, R. T. (1994). Inductive effects on the structure of proline residues. *Int. J. Pept. Protein Res.* **44**, 262-269.

Pedone, E., Ren, B., Ladenstein, R., Rossi, M. & Barolucci, S. (2004). Functional properties of the protein disulfide oxidoreductase from the archaeon *Pyrococcus furiosis*. A member of a novel protein family related to protein disulfide-isomerase. *Eur. J. Biochemistry* **271**, 3427-3448.

Pekkala, M., Hieta, R., Bergmann, U., Kivirikko, K. I., Wierenga, R. K. & Myllyharju, J. (2004). The peptide-substrate binding domain of collagen prolyl 4-hydroxylase is a TPR domain with functional aromatic residues. *J. Biol. Chem.* **279**, 52255-52261.

Peterson, J. R., Ora, A., Van, P. N. & Helenius, A. (1995). Transient, lectin-like association of calreticulin with folding intermediates of cellular and viral glycoproteins. *Mol. Biol. Cell.* **6**, 1173-1184.

Pigiet, V. P. & Schuster, B. J. (1986). Thioredoxin-catalyzed refolding of disulfide-containing proteins. *Proc. Natl. Acad. Sci. USA* **83**, 7643-7647.

Pihlajaniemi, T., Helaakoski, T., Tasanen, K., Myllylä, R., Huhtala, M.-L., Koivu, J. & Kivirikko, K. I. (1987). Molecular cloning of the β-subunit of human prolyl 4-hydroxylase. This subunit and protein disulphide isomerase are products of the same gene. *EMBO J.* **6**, 643-649.

Pind, S., Riordan, J. R. & Williams, D. B. (1994). Participation of the endoplasmic reticulum chaperone calnexin (p88, IP90) in the biogenesis of the cystic fibrosis transmembrane conductance regulator. *J. Biol. Chem.* **269**, 12784-12788.

Pirneskoski, A., Klappa, P., Lobell, M., Williamson, R. A., Byrne, L., Alanen, H. I., Salo, K. E. H., Kivirikko, K. I., Freedman, R. B. & Ruddock, L. W. (2004). Molecular characterization of the principal substrate binding site of the ubiquitous folding catalyst protein disulfide isomerase. *J. Biol. Chem.* **279**, 10374-10381.

Plaxco, K. W., Simons, K. T. & Baker, D. (1998). Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.* **277**, 985-994.

Pollard, M. G., Travers, K. J. & Weissman, J. S. (1998). Ero1p: A novel and ubiquitous protein with an essential role in oxidative protein folding in the endoplasmic reticulum. *Mol. Cell* **1**, 171-182.

Price, J. C., Barr, E. W., Glass, T. E., Krebs, C. & Bollinger, J. M. (2003a). Evidence for hydrogen abstraction from C1 of taurine by the high-spin Fe(IV) intermediate detected during oxygen activation by taurine:α-ketoglutarate dioxygenase (TauD). *J. Am. Chem. Soc.* **125**, 13008-13009.

Price, J. C., Barr, E. W., Tirupati, B., Bollinger, J. M. & Krebs, C. (2003b). The first direct characterization of a high-valent iron intermediate in the reaction of an α-ketoglutarate-dependent dioxygenase: a high-spin Fe(IV) complex in taurine/α-ketoglutarate dioxygenase (TauD) from *Escherichia coli*. *Biochemistry* **42**, 7497-7508.

Prinz, W. A., Åslund, F., Holmgren, A. & Beckwith, J. (1997). The role of the thioredoxin and glutaredoxin pathways in reducing protein disulfide bonds in the *Escherichia coli* cytoplasm. *J. Biol. Chem.* **272**, 15661-15667.

Prockop, D. J., Berg, R. A., Kivirikko, K. I. & Uitto, J. (1976). Intracellular steps in the biosynthesis of collagen. In *Biochemistry of Collagen* (Ramachandran, G. N. & Reddi, A. H., eds.), pp. 163-273. Plenum Press, New York.

Prockop, D. J. & 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. & Kivirikko, K. I. (1995). Collagens: Molecular biology, diseases, and potentials for therapy. *Annu. Rev. Biochemistry* **64**, 403-434.

Prockop, D. J., Sieron, A. L. & Li, S. W. (1998). Procollagen N-proteinase and procollagen C-proteinase. Two unusual metalloproteinases that are essential for procollagen processing probably have important roles in development and cell signalling. *Matrix Biol.* **16**, 399-408.

Proshlyakov, D. A., Henshaw, T. F., Monterosso, G. R., Ryle, M. J. & Hausinger, R. P. (2004). Direct detection of oxygen intermediates in the non-heme Fe enzyme taurine/α-ketoglutarate dioxygenase. *J. Am. Chem. Soc.* **126**, 1022-1023.

Puig, A. & Gilbert, H. F. (1994). Protein disulfide isomerase exhibits chaperone and anti-chaperone activity in the oxidative refolding of lysozyme. *J. Biol. Chem.* **269**, 7764-7771.

Rabbani, L. D., Pagnozzi, M., Chang, P. & Breslow, E. (1982). Partial digestion of neurophysins with proteolytic enzymes: Unusual interactions between bovine neurophysin II and chymotrypsin. *Biochemistry* **21**, 817-826.

Ramshaw, J. A., Werkmeister, J. A. & Glattauer, V. (1996). Collagen-based biomaterials. *Biotechnol. Genet. Eng. Rev.* **13**, 335-382.

Ramshaw, J. A. M., Shah, N. K. & Brodsky, B. (1998). Gly–X–Y tripeptide frequencies in collagen: A context for host–guest triple-helical peptides. *J. Struct. Biol.* **122**, 86-91.

Rapaka, R. S., Renugopalakrishnan, V., Urry, D. W. & Bhatnagar, R. S. (1978). Hydroxylation of proline in polytripeptide models of collagen: stereochemistry of polytripeptide-prolyl hydroxylase interaction. *Biochemistry* **17**, 2892-2898.

Reitch, A., Belin, D., Martin, N. & Beckwith, J. (1996). An in vivo pathway for disulfide bond isomerization in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **93**, 13048-13053.

Ren, B., Tibbelin, G., Pascale, D. d., Rossi, M., Bartolucci, S. & Ladenstein, R. (1998). A protein disulfide oxidoreductase from the archaeon *Pyrococcus furiosus* contains two thioredoxin fold units. *Nature Struct. Biol.* **5**, 602-611.

Reuben, D. M. E. & Bruice, T. C. (1976). Reaction of thiol anions with benzene oxide and malachite green. *J. Am. Chem. Soc.* **98**, 114-121.

Rhoads, R. E. & Udenfriend, S. (1968). Decarboxylation of α-ketoglutarate coupled to collagen proline hydroxylase. *Proc. Natl. Acad. Sci. USA* **60**, 1473-1478.

Rich, A. & Crick, F. H. C. (1961). The molecular structure of collagen. *J. Mol. Biol.* **3**, 483-506.

Riggs-Gelasco, P. J., Price, J. C., Guyer, R. B., Brehm, J. H., Barr, E. W., Bollinger, J. M. & Krebs, C. (2004). EXAFS spectroscopic evidence for an Fe=O unit in the Fe(IV) intermediate observed during oxygen activation by taurine:α-ketoglutarate dioxygenase. *J. Am. Chem. Soc.* **126**, 8108-8109.

Robinson, A. S., Hines, V. & Wittrup, K. D. (1994). Protein disulfide isomerase overexpression increases secretion of foreign proteins in *Saccharomyces cerevisiae*. *BioTechnology* **12**, 381-384.

Ross, C. A. & Poirier, M. A. (2004). Protein aggregation and neurodegenerative disease. *Nat. Med.* **10**, S10-S17.

Roth, W. & Heidemann, E. (1980). Triple helix – coil transition of covalently bridged collagen-like peptides. *Biopolymers* **19**, 1909-1917.

Ruddock, L. W., Hirst, T. R. & Freedman, R. B. (1996). pH-dependence of the dithiol-oxidizing activity of DsbA (a periplasmic protein thiol:disulphide oxidoreductase) and protein disulphide-isomerase: studies with a novel simple peptide substrate. *Biochemistry J.* **315**, 1001-1005.

Rüdiger, S., Germeroth, L., Schneider-Mergener, J. & Bukau, B. (1997). Substrate specificity of the DnaK chaperone determined by screening cellulose-bound peptide libraries. *EMBO J.* **16**, 1501-1507.

Rutkowski, D. T. & Kaufman, R. J. (2004). A trip to the ER: Coping with stress. *Trends Cell. Biol.* **14**, 20-28.

Ryle, M. J., Liu, A., Muthukumaran, R. B., Ho, R. Y. N., Koehntop, K. D., McCracken, J., Que, L. J. & Hausinger, R. P. (2003). O_2- and α-ketoglutarate-dependent tyrosyl radical formation in TauD, an α-keto acid-dependent non-heme iron dioxygenase. *Biochemistry* **42**, 1854-1862.

Ryle, M. J., Padmakumar, R. & Hausinger, R. P. (1999). Stopped-flow kinetic analysis of *Escherichia coli* taurine/α-ketoglutarate dioxygenase: Interactions with α-ketoglutarate, taurine, and oxygen. *Biochemistry* **38**, 15278-15286.

Sachs, A. B. & Varani, G. (2000). Eukaryotic translation initiation: There are (at least) two sides to every story. *Nat. Struct. Biol.* **7**, 356-361.

Sakakibara, S., Inouye, K., Shudo, K., Kishida, Y., Kobayashi, Y. & Prockop, D. J. (1973). Synthesis of $(\text{Pro–Hyp–Gly})_n$ of defined molecular weights. Evidence for the stabilization of collagen triple helix by hydroxyproline. *Biochim. Biophys. Acta* **303**, 198-202.

Sambrook, J. & Russell, D. (2001). *Molecular Cloning: A Laboratory Manual.* 3 edit, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Scheraga, H. A., Wedemeyer, W. J. & Welker, E. (2001). Bovine pancreatic ribonuclease A: Oxidative and conformational folding studies. *Methods Enzymol.* **341**, 189-221.

Scherens, B., Dubois, E. & Messenguy, F. (1991). Determination of the sequence of the yeast *YCL313* gene localized on chromosome III. Homology with the protein disulfide isomerase (PDI gene product) of other organisms. *Yeast* **7**, 185-193.

Schofield, C. J. & Zhang, Z. (1999). Structural and mechanistic studies on 2-oxoglutarate-dependent oxygenases and related enzymes. *Curr. Opin. Struct. Biol.* **9**, 722-731.

Schultz, L. W., Chivers, P. T. & Raines, R. T. (1999). The CXXC motif: Crystal structure of an active-site variant of *Escherichia coli* thioredoxin. *Acta Crystallogr.* **D55**, 1533-1538.

Schwaller, M., Wilkinson, B. & Gilbert, H. F. (2003). Reduction/reoxidation cycles contribute to catalysis of disulfide isomerization by protein disulfide isomerase. *J. Biol. Chem.* **278**, 7154-7159.

Sela, M., White, F. H. & Anfinsen, C. B. (1957). Reductive cleavage of disulfide bridges in ribonuclease. *Science* **125**, 691-692.

Shah, N. K., Ramshaw, J. A. M., Kirkpatrick, A., Shah, C. & Brodsky, B. (1996). A host – guest set of triple-helical peptides: Stability of Gly – X – Y triplets containing common nonpolar residues. *Biochemistry* **35**, 10262-10268.

Sieber, P., Kamber, B., Riniker, B. & Rittel, W. (1980). Iodine oxidation of S-trityl-cysteine-peptides and S-acetamidomethyl-cysteine-peptides containing tryptophan - conditions leading to the formation of tryptophan-2-thioethers. *Helvet. Chim. Acta* **63**, 2358-2363.

Siedler, F., Rudolph-Böhner, S., Doi, M., Musiol, H.-J. & Moroder, L. (1993). Redox potentials of active-site bis(cysteinyl) fragments of thiol-protein oxidoreductases. *Biochemistry* **32**, 7488-7495.

Snellman, A., Keränen, M.-R., Hägg, P. O., Lamberg, A., Hiltunen, K., Kivirikko, K. I. & Pihlajaniemi, T. (2000). Type XIII collagen forms homotrimers with three triple helical collagenous domains and its association into disulfide-bonded trimers is enhanced by prolyl 4-hydroxylase. *J. Biol. Chem.* **275**, 8936-8944.

Solovyov, A., Xiao, R. & Gilbert, H. F. (2004). Sulfhydryl oxidation, not disulfide isomerization, is the principal function of protein disulfide isomerase in yeast *Saccharomyces cerevisiae*. *J. Biol. Chem.* **279**, 34095-34100.

Swartz, J. R. (2001). Advances in *Escherichia coli* production of therapeutic proteins. *Curr. Opin. Biotechnol.* **12**, 195-201.

Szajewski, R. P. & Whitesides, G. M. (1980). Rate constants and equilibrium constants for thiol-disulfide interchange reactions involving oxidized glutathione. *J. Am. Chem. Soc.* **102**, 2011-2026.

Tachibana, C. & Stevens, T. H. (1992). The yeast *EUG1* gene encodes an endoplasmic reticulum protein that is functionally related to protein disulfide isomerase. *Mol. Cell. Biol.* **12**, 4601-4611.

Tachikawa, H., Miura, T., Katakura, Y. & Mizunaga, T. (1991). Molecular structure of a yeast gene, *PDI1*, encoding protein disulfide isomerase that is essential for cell growth. *J. Biochemistry* **110**, 306-313.

Tamamura, H., Matsumoto, F., Sakano, K., Otaka, A., Ibuka, T. & Fujii, N. (1998). Unambiguous synthesis of stromal cell-derived factor-1 by regioselective disulfide bond formation using a DMSO-aqueous HCl system. *Chem. Commun.* **1998**, 151-152.

Tamamura, H., Otaka, A., Nakamura, J., Okubo, K., Koide, T., Ikeda, K. & Fujii, N. (1993). Disulfide bond formation in S-acetamidomethyl cysteine-containing peptides by the combination of silver trifluoromethanesulfonate and dimethylsulfoxide/aqueous HCl. *Tetrahedron Lett.* **34**, 4931-4934.

Tamamura, H., Otaka, A., Nakamura, J., Okubo, K., Koide, T., Ikeda, K., Ibuka, T. & Fujii, N. (1995). Disulfide bond-forming reaction using a dimethyl sulfoxide/aqueous HCl system and its application to regioselective two disulfide bond formation. *Int. J. Peptide Protein Res.* **45**, 312-319.

Tandon, M., Wu, M. & Begley, T. P. (1998). Substrate specificity of human prolyl-4-hydroxylase. *Bioorg. Med. Chem. Lett.* **8**, 1139-1144.

Thomas, P. J., Ko, Y. H. & Pedersen, P. L. (1992a). Altered protein folding may be the molecular basis of most cases of cystic fibrosis. *FEBS Lett.* **312**, 7-9.

Thomas, P. J., Qu, B.-H. & Pederson, P. L. (1995). Defective protein folding as a basis of human disease. *Trends Biochemistry Sci.* **20**, 456-459.

Thomas, P. J., Shenbagamurthi, P., Sondek, J., Hullihen, J. M. & Pedersen, P. L. (1992b). The cystic fibrosis transmembrane conductance regulator. Effects of the most common cystic fibrosis-causing mutation on the secondary structure and stability of a synthetic peptide. *J. Biol. Chem.* **267**, 5727-5730.

Thornburg, L. D., Lai, M.-T., Wishnok, J. S. & Stubbe, J. (1993). A non-heme iron protein with heme tendencies: an investigation of the substrate specificity of thymine hydroxylase. *Biochemistry* **32**, 14023-14033.

Trojanowska, M., LeRoy, E. C., Eckes, B. & Krieg, T. (1998). Pathogenesis of fibrosis: type 1 collagen and the skin. *J. Mol. Med.* **76**, 266-274.

Tsai, B. & Rapoport, T. A. (2002). Unfolded cholera toxin is transferred to the ER membrane and released from protein disulfide isomerase upon oxidation by Ero1. *J. Cell. Biol.* **159**, 207-215.

Tu, B. P., Ho-Schleyer, S. C., Travers, K. J. & Weissman, J. S. (2000). Biochemical basis of oxidative protein folding in the endoplasmic reticulum. *Science* **290**, 1571-1574.

Tu, B. P. & Weissman, J. S. (2002). The FAD- and O_2-dependent reaction cycle of Ero1-mediated oxidative protein folding in the endoplasmic reticulum. *Mol. Cell* **10**, 983-994.

Tuderman, L., Kuutti, E.-R. & Kivirikko, K. I. (1975). An affinity-column procedure using poly(L-proline) for the purification of prolyl hydroxylase. *Eur. J. Biochemistry* **52**, 9-16.

Tuderman, L., Myllyla, R. & Kivirikko, K. I. (1977). Mechanism of the prolyl hydroxylase reaction. 1. Role of co-substrates. *Eur. J. Biochemistry* **80**, 341-348.

Valegård, K., Terwisscha 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. & Andersson, I. (1998). Structure of a cephalosporin synthase. *Nature* **394**, 805-809.

Vitoux, B., Aubry, A., Cung, M. T. & Marraud, M. (1986). N-methyl peptides. 7. Conformational perturbations induced by N-methylation of model dipeptides. *Int. J. Pept. Protein Res.* **27**, 617-632.

von der Haar, T., Gross, J. D., Wagner, G. & McCarthy, J. E. G. (2004). The mRNA cap-binding protein eIF4E in post-transcriptional gene expression. *Nat. Struct. Biol.* **11**, 503-511.

Vuorela, A., Myllyharju, J., Nissi, R., Pihlajaniemi, T. & Kivirikko, K. I. (1997). Assembly of human prolyl 4-hydroxylase and type III collagen in the yeast *Pichia pastoris*: Formation of a stable enzyme tetramer requires coexpression with collagen and assembly of a stable collagen requires coexpression with prolyl 4-hydroxylase. *EMBO J.* **16**, 6702-6712.

Vuori, K., Myllylä, R., Pihlajaniemi, T. & Kivirikko, K. I. (1992a). Expression and site-directed mutagenesis of human protein disulfide isomerase in *Escherichia coli*. *J. Biol. Chem.* **267**, 7211-7214.

Vuori, K., Pihlajaniemi, T., Marttila, M. & 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., Myllylä, R. & 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. & Gilbert, H. F. (1997). Scanning and escape during protein-disulfide isomerase-assisted protein folding. *J. Biol. Chem.* **272**, 8845-8848.

Walker, K. W., Lyles, M. M. & Gilbert, H. F. (1996). Catalysis of oxidative protein folding by mutants of protein disulfide isomerase with a single active-site cysteine. *Biochemistry* **35**, 1972-1980.

Walter, S. & Buchner, J. (2002). Molecular chaperones - Cellular machines for protein folding. *Angew. Chem. Int. Ed. Engl.* **41**, 1098-1113.

Wang, S., Sakai, H. & Wiedmann, M. (1995). NAC covers ribosome-associated nascent chains thereby forming a protective environment for regions of nascent chains just emerging from the peptidyl transferase center. *J. Cell Biol.* **130**, 519-528.

Weissman, J. S. & Kim, P. S. (1993). Efficient catalysis of disulphide bond rearrangements by protein disulphide isomerase. *Nature* **365**, 185-188.

Werkmeister, J. A. & Ramshaw, J. A. M., Eds. (1992). *Collagen Biomaterials*. Barking, Essex, England: Elsevier Science.

Westphal, V., Spetzler, J. C., Meldal, M., Christensen, U. & Winther, J. R. (1998). Kinetic analysis of the mechanism and specificity of protein-disulfide isomerase using fluorescence-quenched peptides. *J. Biol. Chem.* **273**, 24992-24999.

Wetterau, J. R., Combs, K. A., Spinner, S. N. & Joiner, B. J. (1990). Protein disulfide isomerase is a component of the microsomal triglyceride transfer protein complex. *J. Biol. Chem.* **265**, 9800-9807.

Whitesides, G. M., Lilburn, J. E. & Szajewski, R. P. (1977). Rates of thiol-disulfide interchange reactions between mono- and dithiols and Ellman's reagent. *J. Org. Chem.* **42**, 332-338.

Wilkinson, B. M., Regnacq, M. & Stirling, C. J. (1997). Protein translocation across the membrane of the endoplasmic reticulum. *J. Membrane Biol.* **155**, 189-197.

Wilmouth, R. C., Turnbull, J. J., Welford, R. W. D., Clifton, I. J., Prescott, A. G. & Schofield, C. J. (2002). Structure and mechanism of anthocyanidin synthase from *Arabidopsis thaliana*. *Structure* **10**, 93-103.

Wilson, J. M., Bayer, R. J. & Hupe, D. J. (1977). Structure-reactivity correlations for the thiol-disulfide interchange reaction. *J. Am. Chem. Soc.* **99**, 7922-7926.

Winter, A. D. & 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.

Winter, J., Lilie, H. & Rudolph, R. (2002). Recombinant expression and in vitro folding of proinsulin are stimulated by the synthetic dithiol Vectrase-P. *FEMS Microbiol. Lett.* **213**, 225.

Wittrup, K. D. (1995). Disulfide bond formation and eucaryotic secretory productivity. *Curr. Opin. Biotechnol.* **6**, 203-208.

Woycechowsky, K. J. & Raines, R. T. (2000). Native disulfide bond formation in proteins. *Curr. Opin. Chem. Biol.* **4**, 533-539.

Woycechowsky, K. J. & Raines, R. T. (2003). The CXC motif: A functional mimic of protein disulfide isomerase. *Biochemistry* **42**, 5387-5394.

Woycechowsky, K. J., Wittrup, K. D. & Raines, R. T. (1999). A small-molecule catalyst of protein folding *in vitro* and *in vivo*. *Chem. Biol.* **6**, 871-879.

Wu, M., Moon, H.-S. & Begley, T. P. (1999). Mechanism-based inactivation of the human prolyl-4-hydroxylase by 5-oxaproline-containing peptides: Evidence for a prolyl radical intermediate. *J. Am. Chem. Soc.* **121**, 587-588.

Wu, M., Moon, H.-S., Pirskanen, A., Myllyharju, J., Kivirikko, K. I. & Begley, T. P. (2000). Mechanistic studies on prolyl 4-hydroxylase: The vitamin C requiring uncoupled oxidation. *Bioorg. Med. Chem. Lett.* **10**, 1511-1514.

Xiao, R., Solovyov, A., Gilbert, H. F., Holmgren, A. & Lundström-Ljung, J. (2001). Combinations of protein-disulfide isomerase domains show that there is little correlation between isomerase activity and wild-type growth. *J. Biol. Chem.* **276**, 27975-27980.

Xiao, R., Wilkinson, B., Solovyov, A., Winther, J. R., Holmgren, A., Lundström-Ljung, J. & Gilbert, H. F. (2004). The contributions of protein disulfide isomerse and its homologs to oxidative protein folding in the yeast ER. *J. Biol. Chem.* **279**, 49780-49786.

Yang, C., Hillas, P. J., Báez, J. A., Nokelainen, M., Balan, J., Tang, J., Spiro, R. & Polarek, J. W. (2004). The application of recombinant human collagen in tissue engineering. *Biodrugs* **18**, 103-119.

Yang, Y., Janich, S., Cohn, J. A. & Wilson, J. M. (1993). The common variant of cystic fibrosis transmembrane conductance regulator is recognized by hsp70 and degraded in a pre-golgi nonlysosomal compartment. *Proc. Natl. Acad. Sci. USA* **90**, 9480-9484.

Zagari, A., Nemethy, G. & Scheraga, H. A. (1990). The effect of the L-azetidine-2-carboxylic acid residue on protein conformation. III. Collagen-like poly(tripeptide)s. *Biopolymers* **30**, 967-974.

Zapun, A., Bardwell, J. C. & Creighton, T. E. (1993). The reactive and destabilizing disulfide bond of DsbA, a protein required for protein disulfide bond formation in vivo. *Biochemistry* **32**, 5083-5092.

Zapun, A., Jakob, C. A., Thomas, D. Y. & Bergeron, J. J. (1999). Protein folding in a specialized compartment: The endoplasmic reticulum. *Structure Fold. Des.* **7**, R173-182.

Zapun, A., Missiakas, D., Raina, S. & Creighton, T. E. (1995). Structural and functional characterization of DsbC, a protein involved in disulfide bond formation in *Escherichia coli*. *Biochemistry* **34**, 5075-5089.

Zhang, K. & Kaufman, R. J. (2004). Signaling the unfolded protein response from the endoplasmic reticulum. *J. Biol. Chem.* **279**, 25935-25938.

Zhang, R. & 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, Z., Ren, J., Stammers, D. K., Baldwin, J. E., Harlos, K. & Schofield, C. J. (2000). Structural origins of the selectivity of the trifunctional oxygenase clavaminic acid synthase. *Nat. Struct. Biol.* **7**, 127-133.

Zhou, J., Gunsior, M., Bachmann, B. O., Townsend, C. A. & Solomon, E. I. (1998). Substrate binding to the α-ketoglutarate-dependent non-heme iron enzyme clavaminate synthase 2: Coupling mechanism of oxidative decarboxylation and hydroxylation. *J. Am. Chem. Soc.* **120**, 13539-13540.

Zhu, X., Zhao, X., Burkholder, W. F., Gragerov, A., Ogata, C. M., Gottesman, M. E. & Hendrickson, W. A. (1996). Structural analysis of substrate binding by the molecular chaperone DnaK. *Science* **272**, 1606-1614.