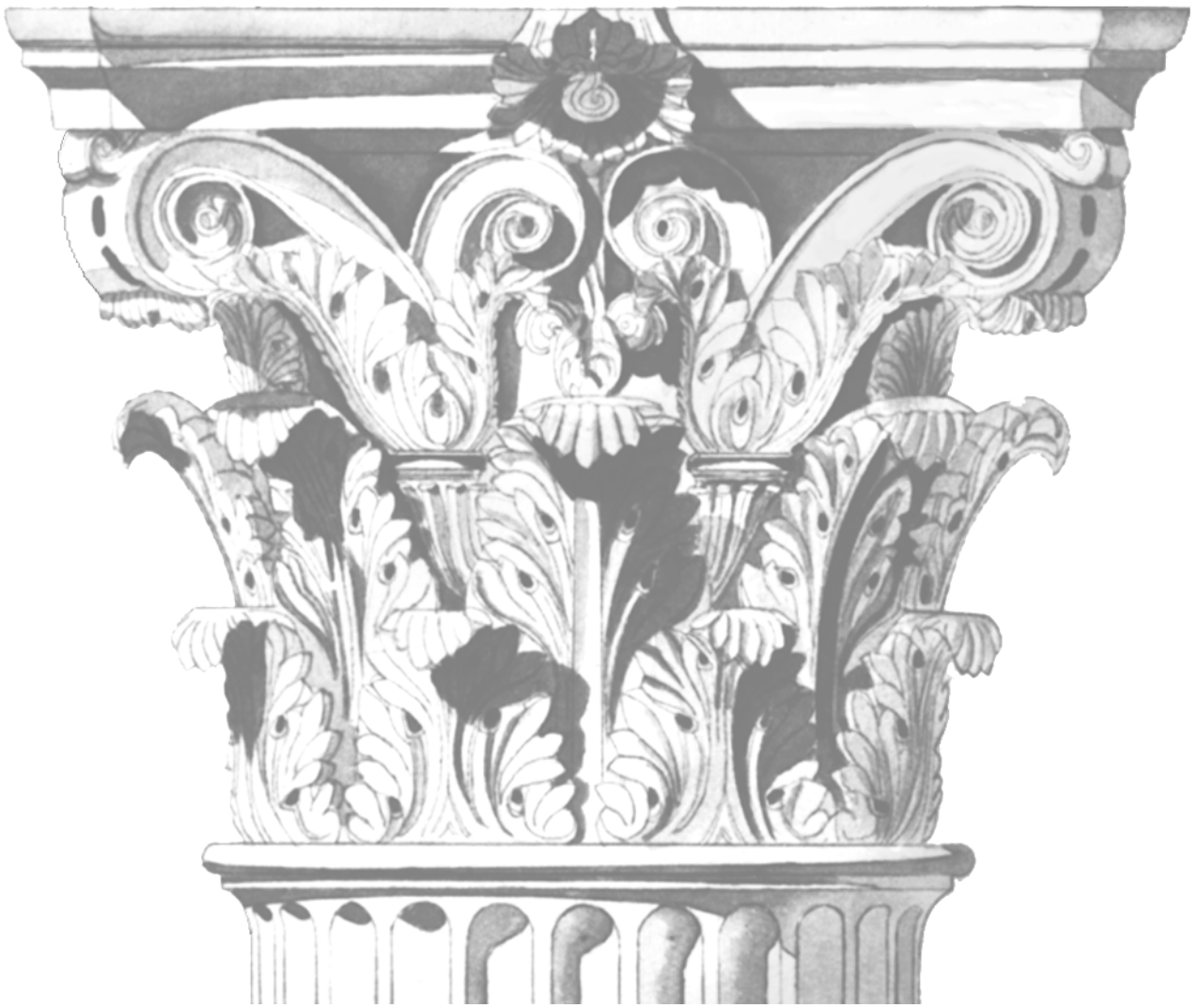
COLLAGEN PROLYL 4-HYDROXYLASE

Characterization of a novel vertebrate isoenzyme and the main *Caenorhabditis elegans* enzyme forms, and effect of inactivation of one of the two catalytic sites in the enzyme tetramer

LIISA KUKKOLA

Collagen Research Unit,
Biocenter Oulu and
Department of Medical Biochemistry
and Molecular Biology,
University of Oulu

OULU 2003



LIISA KUKKOLA

COLLAGEN PROLYL 4-HYDROXYLASE

Characterization of a novel vertebrate isoenzyme and the main *Caenorhabditis elegans* enzyme forms, and effect of inactivation of one of the two catalytic sites in the enzyme tetramer

Academic Dissertation to be presented with the assent of the Faculty of Medicine, University of Oulu, for public discussion in the Auditorium L101 of the Department of Medical Biochemistry and Molecular Biology, on December 5th, 2003, at 10 a.m.

OULUN YLIOPISTO, OULU 2003

Supervised by
Docent Johanna Myllyharju
Professor Kari Kivirikko

Reviewed by
Professor Leena Brückner-Tuderman
Docent Merja Perälä

ISBN 951-42-7203-X (URL: http://herkules.oulu.fi/isbn951427203X/)

ALSO AVAILABLE IN PRINTED FORMAT
Acta Univ. Oul. D 764, 2003
ISBN 951-42-7202-1
ISSN 0355-3221 (URL: http://herkules.oulu.fi/issn03553221/)

OULU UNIVERSITY PRESS
OULU 2003

Kukkola, Liisa, Collagen prolyl 4-hydroxylase. Characterization of a novel vertebrate isoenzyme and the main *Caenorhabditis elegans* enzyme forms, and effect of inactivation of one of the two catalytic sites in the enzyme tetramer

Collagen Research Unit, Biocenter Oulu, and Department of Medical Biochemistry and Molecular Biology, University of Oulu, P.O.Box 5000, FIN-90014 University of Oulu, Finland
Oulu, Finland
2003

Abstract

Collagen prolyl 4-hydroxylases catalyze the hydroxylation of proline residues in collagens. The vertebrate enzymes are $\alpha_2\beta_2$ tetramers in which the β subunit is identical to protein disulphide isomerase (PDI). Two isoforms of the catalytic α subunit have been identified in vertebrates, forming type I $[\alpha(I)]_2\beta_2$ and type II $[\alpha(II)]_2\beta_2$ collagen prolyl 4-hydroxylase tetramers.

This thesis reports on the cloning and characterization of a third vertebrate α subunit isoform, α(III). The recombinant human α(III) isoform associates with PDI to form an active type III collagen prolyl 4-hydroxylase tetramer, and its K_m values for the cosubstrates are very similar to those of the type I and II enzymes, those for a peptide substrate and an inhibitor being found to lie between the two. The α(III) mRNA is expressed in all tissues studied but at much lower levels than the α(I) mRNA.

A novel mixed tetramer PHY-1/PHY-2/$(PDI\text{-}2)_2$ was found to be the main collagen prolyl 4-hydroxylase form produced in the nematode *Caenorhabditis elegans in vivo* and *in vitro*. However, mutant nematodes can compensate for the lack of the mixed tetramer by increasing the assembly of PHY-1/PDI-2 and PHY-2/PDI-2 dimers, these forms also being unique. The catalytic properties of the recombinant mixed tetramer were characterized, and it was shown by the analysis of mutant worms that PHY-1 and PHY-2 represent the only catalytic subunits needed for the hydroxylation of cuticular collagens.

The roles of the two catalytic sites in a collagen prolyl 4-hydroxylase tetramer were studied by using the *C. elegans* mixed tetramer and a hybrid *C. elegans* PHY-1/human PDI dimer. An increase in the chain length of the peptide substrate led to an identical decrease in the K_m values in both enzyme forms. It is thus clear that two catalytic sites are not required for efficient hydroxylation of long peptides, and their low K_m values most probably result from more effective binding to the peptide-substrate-binding domain. Inactivation of one catalytic site in the mixed tetramer reduced the activity by more than 50%, indicating that the remaining wild-type subunit cannot function fully independently.

Keywords: C. elegans, collagen, isoenzyme, prolyl 4-hydroxylase

Acknowledgements

This work was carried out at the Collagen Research Unit and the Department of Medical Biochemistry and Molecular Biology, University of Oulu, and Biocenter Oulu during the years 1997-2003.

I want to express my deepest gratitude to my supervisor, Docent Johanna Myllyharju, for guiding me into the scientific world and for her constant support and everlasting optimism during the last six years. I am very grateful to Professor Kari Kivirikko for providing me with the opportunity to work in his research group. His vast knowledge of collagens and collagen modifying enzymes has formed the basis for this research. I would also like to thank Professor Taina Pihlajaniemi for providing excellent research facilities in the Department, and Professor Leena Ala-Kokko for her optimism and encouragement. I also wish to express my respect to Professor Ilmo Hassinen for his distinguished scientific work. Professor Peppi Karppinen deserves my warmest thanks for her friendship, for her excellent guidance during the last *C. elegans* project and for encouraging me in my medical career. Her positive attitude and interest in work and life in general is admirable.

I am grateful to Professor Leena Brückner-Tuderman and Docent Merja Perälä for their careful reviewing of this thesis and their valuable comments on it. I also wish to thank Malcolm Hicks, M.A., for revising the English language of this thesis and the original articles.

I wish to express my sincere thanks to Dr. Antony Page and Dr. Alan Winter for their practical and skilful guidance in the *C. elegans* projects and for offering me an opportunity to visit their laboratory in Glasgow, Scotland.

I would like to thank the staff of the Department, especially Auli Kinnunen, Marja-Leena Kivelä, Seppo Lähdesmäki, Pertti Vuokila and Marja-Leena Karjalainen, for their kind and helpful attitude in solving all manner of problems. The computer support crew is also acknowledged for quick and professional help. Kindest thanks go to Merja Nissilä, Anu Myllymäki and Outi Mänty for their excellent, skilful technical assistance and interest in everyday work in the lab.

My warmest thanks belong to the three wonderful young women of our office ‘Paradise’, Reija Hieta, Outi Pakkanen and Hanna-Leena Ruuska, and the two former members of the ‘Paradise’ Peppi Karppinen and Annamari Ruddock. You have brought

sunshine to many cold, rainy days, and it has been a real privilege to share an office with you. I also wish to thank all the people working in our group for the fruitful and relaxed atmosphere. Katriina Keskiaho especially deserves my deepest thanks for her friendship and unforgettable, sometimes even desperate, moments with the worms. I would also like to thank Mirka Vuoristo for cheerful moments when organizing the dissertation party.

I owe my deepest gratitude to my parents Helena and Antti and my brother Juha for their love and encouragement. My mother deserves my sincere thanks for supporting my studies with a true interest since the very beginning of my school years. I also wish to thank Katja Kykyri for being a faithful friend and such excellent company for so many years. Above all, I owe my greatest gratitude to Jaakko Erkkilä, whose love, confidence and support has been essential in bringing my thesis to its completion. I am grateful to him for sharing interesting life outside research work.

This work was supported by the Finnish Medical Foundation Duodecim and the Finnish Centre of Excellence Programme (2000-2005) of the Academy of Finland.

Oulu, October 2003 Liisa Kukkola

Abbreviations

bp(s)	base pair(s)
cDNA	complementary DNA
C. elegans	*Caenorhabditis elegans*
C-P4H	Collagen prolyl 4-hydroxylase
CSM	Collagen-related structural motif
dpy-18	*C. elegans phy-1* knock-out strain
dsRNAi	double stranded RNA interference
ER	Endoplasmic reticulum
EST	Expressed sequence tag
FIH	Factor inhibiting HIF
HIF	Hypoxia-inducible factor
HIF-P4H	HIF prolyl 4-hydroxylase
IPNS	Isopenicillin *N* synthase
K_m	Michaelis-Menten constant
mRNA	Messenger RNA
NMR	Nuclear magnetic resonance
P4H	Prolyl 4-hydroxylase
PBCV-1	*Paramecium bursaria Chlorella* virus-1
PCR	Polymerase chain reaction
PDI	Protein disulphide isomerase
PHY	*C. elegans* collagen prolyl 4-hydroxylase α subunit
phy	A gene encoding PHY
RACE	Rapid amplification of cDNA ends
RNAi	RNA interference
VHL	Von Hippel-Lindau
X (in -Gly-X-Y-)	any amino acid
Y (in -Gly-X-Y-)	any amino acid

List of original articles

This thesis is based on the following articles, which are referred to in the text by their Roman numerals:

I Kukkola L, Hieta R, Kivirikko KI & Myllyharju J (2003) Identification and characterization of a third human, rat and mouse collagen prolyl 4-hydroxylase isoenzyme. J Biol Chem, in press.

II Myllyharju J, Kukkola L, Winter AD & Page AP (2002) The exoskeleton collagens in *Caenorhabditis elegans* are modified by prolyl 4-hydroxylases with unique combinations of subunits. J Biol Chem 277: 29187-29196.

III Kukkola L, Koivunen P, Pakkanen O, Page AP & Myllyharju J (2003) Mutation of one of the two catalytic sites in a collagen prolyl 4-hydroxylase tetramer inactivates the enzyme by more than half. Manuscript.

Contents

1 Introduction

Collagens form a superfamily of structural proteins of the extracellular matrix. They are expressed in all tissues of the human body and are involved in multiple important functions supporting the architecture, strength and development of tissues and affecting cell attachment, proliferation, migration and differentiation.

Collagen prolyl 4-hydroxylases (C-P4Hs) have an essential role in collagen synthesis, as the hydroxylation of proline residues in the α chains is required for the formation of stable collagen trimers. In vertebrates, the C-P4Hs are $\alpha_2\beta_2$ tetramers, in which protein disulphide isomerase (PDI) serves as the β subunit. Two α subunit isoforms, forming type I $[\alpha(I)_2]\beta_2$ and type II $[\alpha(II)_2]\beta_2$ C-P4Hs, have been cloned and characterized from human and mouse. In addition to vertebrates, prolyl 4-hydroxylases (P4H) have been identified and characterized from fruit fly, nematode, plant, algal and viral sources. Plants do not express collagens, but P4H is needed in the hydroxylation of many of their cell-wall glycoproteins. The plant and viral P4Hs have been shown to be monomers, while a C-P4H characterized from the fruit fly *Drosophila melanogaster* resembles vertebrate C-P4Hs in being an $\alpha_2\beta_2$ tetramer, the filarial nematode *Brugia malayi* C-P4H being an α_4 homotetramer. The genomes of *Arabidopsis thaliana, B. malayi, Caenorhabditis elegans* and *D. melanogaster* have been fully or almost fully sequenced, and their analysis has revealed several genes encoding C-P4H α subunit-like polypeptides.

The aim of the present research was to clone and characterize a third C-P4H α subunit isoform from human, rat and mouse sources, to determine its expression pattern in various human tissues and to express an active recombinant human type III C-P4H for catalytic studies.

The present thesis also describes the identification and characterization of three unique *C. elegans* C-P4H forms responsible for the hydroxylation of cuticle collagens. The *C. elegans* mixed tetramer PHY-1/PHY-2/$(PDI\text{-}2)_2$ and a PHY-1/PDI dimer were used further to study the roles of the two catalytic sites in a C-P4H tetramer.

2 Review of the literature

2.1 4-Hydroxyproline in proteins

The majority of 4-hydroxyproline residues in animals are found in collagens and more than 20 other proteins that have collagen-like domains but are not defined as collagens (for reviews, see Kivirikko & Pihlajaniemi 1998, Myllyharju & Kivirikko 2001, 2003, Kielty & Grant 2002). Collagen molecules consist of three polypeptide chains, called α chains. Each α chain has a left-handed helix and the three chains are coiled into a right-handed triple helix around each other. The triple-helical regions of the α chains typically consist of repeating -Gly-X-Y- sequences. The occurrence of glycine in every third position in each chain has a critical role, since, being the smallest amino acid, glycine can fit into the limited space in the centre of the triple helix, where the three α chains come together (see Jenkins & Raines 2002, Kielty & Grant 2002, Myllyharju & Kivirikko 2003). The proline residues frequently found in the X positions stabilize the triple helix, as their ring structure prevents rotation around the N-C_α bond in the polypeptide chains, rotation around the C_α-C=O bond also being restricted (see Jenkins & Raines 2002, Kielty & Grant 2002). The 4-hydroxyproline residues frequently found in the Y positions have a critical role in providing the collagen triple helices with thermal stability. Non-hydroxylated type I collagen is denatured at 24°C while a triple helix consisting of hydroxylated α chains is stable up to 39°C (Berg & Prockop 1973, Rosenbloom *et al*. 1973). The 4-hydroxyproline content of the most abundant collagen, type I, is about 100 amino acids per 1000, small but distinct differences in 4-hydroxyproline content being found between the various collagen types (Kivirikko *et al*. 1992). The collagen prolyl 4-hydroxylases (C-P4Hs) have a central role in collagen synthesis, as they catalyze the hydroxylation of peptidyl proline residues in the Y positions of the repeating -Gly-X-Y- triplets that are characteristic of collagens and collagen-like domains (see Kivirikko & Myllyharju 1998, Kivirikko & Pihlajaniemi 1998, Myllyharju 2003).

Several mechanisms have been proposed by which 4-hydroxyproline residues stabilize the triple-helical structure. In 1994 the presence of water molecules bound to the hydroxy groups of 4-hydroxyproline residues was confirmed by X-ray diffraction

analysis, and it was found that the 4-hydroxyprolines most frequently bind two water molecules, thus forming an interchain link to the amide oxygen of another 4-hydroxyproline residue. The network of water bridges was thought to offer a good model for the interpretation of experimental results on collagen stability and assembly (Bella *et al*. 1994, 1995). Replacement of the 4-hydroxyproline residues in a collagen-like polytripeptide $(Pro\text{-}4Hyp\text{-}Gly)_{10}$ with 4(*R*)-fluoro-L-proline residues surprisingly enhanced the triple helix stability, however. In this model the formation of water bridges and hydrogen bonds is prevented, but the hyperstability was thought to arise from the strong inductive effects of the most electronegative atom, fluorine (Holmgren *et al*. 1998, 1999, Ramshaw & Brodsky 2003). 4-Hydroxyproline residues probably act in the same way as fluoroproline and stabilize the collagen triple helix by means of a stereoelectronic effect which fixes the pyrrolidine ring pucker and thus preorganizes all three main-chain torsion angles (Jenkins & Raines 2002).

Collagens and proteins with collagen-like domains form a superfamily in various animal species. At the moment at least 27 collagen types have been identified, with at least 42 different α chains, and the number of the known members of the family is increasing continuously. A collagen trimer may consist of two or three different α chains or three identical chains. Most collagens form supramolecular assemblies, which can be divided into nine distinct families: (1) fibril-forming collagens (types I, II, III, V, XI, XXIV, XXVII); (2) fibril-associated collagens with interrupted triple helices (FACITs) located on the surface of the fibrils, and structurally related collagens (types IX, XII, XIV, XVI, XIX, XX, XXI, XXII, XXVI); (3) the family of collagens VIII and X forming hexagonal networks; (4) the type IV collagens, located in basement membranes; (5) type VI collagen, forming beaded filaments; (6) type VII collagen, forming anchoring fibrils for basement membranes; (7) collagens containing transmembrane domains (types XIII, XVII, XXIII, XXV); and (8) the family of type XV and XVIII collagens. The modes of supramolecular assembly of families 7 and 8 are so far unknown. Family number 9 consists of proteins containing triple-helical collagenous domains but not defined as collagens (for a review, see Myllyharju & Kivirikko 2003).

Elastin is the major component of elastic fibres (for a review, see Rosenbloom & Abrams 2002), and it also has repeating -Gly-X-Y- sequences with 4-hydroxyproline residues in the Y-positions, but no collagen-like triple-helical domains. The 4-hydroxyproline content of elastin usually ranges from 10 to 25 per 1000 but major variations have been observed (Kivirikko *et al*. 1992, Kivirikko & Pihlajaniemi 1998, Rosenbloom & Abrams 2002).

Hydroxyproline is further found in the hypoxia-inducible factor (HIF) in which the hydroxylation of Pro564, and in some cases also of Pro402, in the HIF-1α plays a critical role in the adaptation to hypoxia at the cellular and systemic level (for reviews, see Semenza 2000, 2001, Fedele *et al*. 2002).

Plants do not express collagens, but 4-hydroxyproline is found in many plant cell-wall glycoproteins such as extensins, the proline-rich proteins, the solanaceous lectins and the arabinogalactan proteins (Showalter 1993).

It was previously believed that bacterial and viral proteins did not contain hydroxyproline, but interestingly, a collagen-related structural motif (CSM) has been identified in proteins from bacteria, bacteriophages, and viruses (Medveczky *et al*. 1993, Smith *et al*. 1998, Rasmussen *et al.* 2003). The bacterial and viral CSMs differ

significantly from the vertebrate collagens in their amino acid content and distribution; having a lower proline content and threonine instead of proline as the dominating amino acid in the Y position of the -Gly-X-Y- repeats in some of the CSMs identified. Molecular modelling suggests that a threonine in the Y position could substitute for 4-hydroxyproline in the stabilization of the collagen-like triple helix of bacterial CSMs (Rasmussen *et al.* 2003). No bacterial homologues of eukaryotic prolyl 4-hydroxylases have been found in extensive similarity searches (Rasmussen *et al.* 2003), but a proline 4-hydroxylase that acts on free L-proline has been identified in *Streptomyces* (Lawrence *et al*. 1996).

Evidence for the presence of 4-hydroxyproline in viral proteins was obtained by the identification, cloning and characterization of a prolyl 4-hydroxylase from a eukaryotic algal virus, *Paramecium bursaria Chlorella* virus-1. The recombinant viral enzyme hydroxylated the synthetic peptides $(Pro\text{-}Ala\text{-}Pro\text{-}Lys)_n$, $(Ser\text{-}Pro\text{-}Lys\text{-}Pro\text{-}Pro)_5$ and $(Pro\text{-}Glu\text{-}Pro\text{-}Pro\text{-}Ala)_5$, corresponding to the proline-rich repeats encoded by the viral genome, and also the collagen-like peptide $(Pro\text{-}Pro\text{-}Gly)_{10}$ and poly(L-proline) (Eriksson *et al*. 1999).

2.2 Collagen prolyl 4-hydroxylases and related enzymes

2.2.1 Vertebrate collagen prolyl 4-hydroxylases

Collagen prolyl 4-hydroxylases (EC 1.14.11.2, procollagen-proline, 2-oxoglutarate 4-dioxygenases) are located within the lumen of the endoplasmic reticulum and catalyze the formation of the 4-hydroxy group in peptidyl proline residues.

2.2.1.1 Molecular properties

The vertebrate collagen prolyl 4-hydroxylases (C-P4Hs) are $\alpha_2\beta_2$ tetramers with a molecular weight of 240 000, in which the enzyme and chaperone protein disulphide isomerase (PDI) serves as the β subunit. Two isoforms of the catalytic α subunit have been identified in human and mouse sources forming the type I $[\alpha(I)]_2\beta_2$ and type II $[\alpha(II)]_2\beta_2$ C-P4H enzyme tetramers (Helaakoski *et al*. 1989, 1995, Annunen *et al*. 1997). The two types of α subunit do not appear to form a mixed $\alpha(I)\alpha(II)\beta_2$ C-P4H tetramer, as indicated in data obtained in recombinant coexpression experiments with insect cells (Annunen *et al*. 1997). The human α(I) and α(II) subunits consist of 517 and 514 amino acids, respectively, and a signal sequence of 17 and 21 additional residues. The overall identity between the human α(I) and α(II) amino acid sequences is 64%, whereas the catalytic C-terminal regions are better conserved, their identity being 80% (Kivirikko & Myllyharju 1998, Myllyharju 2003). The human α(I) and α(II) subunits contain five

conserved cysteine residues, the α(II) subunit having an additional cysteine between the fourth and fifth conserved cysteine residues (Helaakoski *et al*. 1995, Annunen *et al*. 1997). Site-directed mutagenesis experiments suggest that an intramolecular disulphide bond is formed between the second and third and between the fourth and fifth cysteines of the α(I) subunit, these disulphide bonds being essential for $\alpha_2\beta_2$ tetramer formation (John & Bulleid 1994, Lamberg *et al*. 1995). The α(I) and α(II) subunits each contain two asparagine residues that are potential N glycosylation sites. However, no effect on tetramer formation or C-P4H activity was observed upon mutation of one or both of these asparagines to glutamine in the recombinant α(I) subunit (Lamberg *et al*. 1995). The type I enzyme is the main C-P4H form in most cell types and tissues studied, type II being the major form in chondrocytes, osteoblasts, endothelial cells and cells of epithelial structures (Annunen *et al*. 1998, Nissi *et al*. 2001).

Many attempts to determine the 3D structure of the C-P4H tetramer have been made over the years, but they all have been unsuccessful. The α subunit itself is an insoluble protein (Vuori *et al.* 1992b) and thus unsuitable for crystallization. Another approach involves identification of the domains of the α subunit and attempts to determine their structures. Limited proteolysis studies performed on the type I C-P4H tetramer bound to poly(L-proline) sepharose showed that the peptide-binding domain of the α(I) subunit is located between residues Gly138 and Ser244 in the N-terminal region and is thus distinct from the C-terminal catalytic region. The recombinant peptide-binding domain produced in *E. coli* is soluble, suitable for structural studies and distinct from the previously described proline-rich peptide-binding domains, and thus represents a new type of proline-rich peptide-binding module (Myllyharju & Kivirikko 1999). According to NMR assignments the domain is composed of five α helices and a short β strand (Hieta *et al*. 2003). The recombinant peptide-binding domain has been crystallized successfully as well (Pekkala *et al*. 2003) and work is in progress on resolving its structure (Pekkala M, Hieta R, Myllyharju J, Kivirikko KI, Wierenga RK, unpublished data).

Protein disulphide isomerase (PDI) is a multifunctional protein located within the lumen of the rough endoplasmic reticulum that catalyzes the formation, breakage and rearrangement of disulphide bonds during protein folding (for a review, see Kivirikko and Myllyharju 1998, Noiva 1999). It serves as the β subunit in all known vertebrate C-P4Hs, but there is evidence that the corresponding P4Hs in lower organisms such as algae (Kaska *et al*. 1987, 1988) and plants (Wojtaszek *et al*. 1999, Hieta & Myllyharju 2002) are α monomers and thus not associated with PDI. PDI also has several other functions, e.g. it acts as the β subunit in the microsomal triglyceride transfer protein dimer (Wetterau *et al*. 1991) and as a chaperone-like polypeptide that binds various polypeptides within the lumen of the endoplasmic reticulum (Noiva *et al*.1993) and assists in their folding (LaMantia & Lennarz 1993, Cai *et al*. 1994, Otsu *et al*. 1994, Lamandé & Bateman 1999).

The PDI polypeptide has been cloned from various organisms: animals, yeasts, micro-organisms and plants (see Kivirikko & Pihlajaniemi 1998). The human PDI polypeptide consists of 491 amino acids and a signal sequence of 17 additional residues (Pihlajaniemi *et al*. 1987), and it is built up of four domains, *a, b, b*' and *a'* together with a C-terminal extension *c* (Edman *et al*. 1985, Pihlajaniemi *et al*. 1987). The *a* and *a*' domains each contain a thioredoxin-like -Cys-Gly-His-Cys- sequence which represents two independently acting catalytic sites of PDI activity (Hawkins & Freedman 1991, Vuori *et*

al. 1992a). Mutation of the two -Cys-Gly-His-Cys- sequences to -Ser-Gly-His-Cys- did not reduce the assembly of the C-P4H tetramer, however, nor its P4H activity (Vuori *et al*. 1992c). NMR analyses have shown that the domains *a* and *a'*, and also domain *b*, although it lacks the conserved catalytic site residues, all have a thioredoxin fold (Kemmink *et al*. 1996, 1997, Dijkstra *et al*. 1999). Preliminary data on the *b'* domain suggest that it has a similar thioredoxin fold, and thus the *b* and *b'* domains represent inactive thioredoxin modules (Kemmink *et al*. 1997). The majority of the amino acids in the extension *c* are acidic, and it has been suggested that this extension may act as a Ca^{2+} binding region (Lebeche *et al*. 1994).

The C terminus of the PDI polypeptide contains a -Lys-Asp-Glu-Leu- motif, which retains the PDI polypeptide within the lumen of the endoplasmic reticulum (Pelham 1990). Deletion of the retention signal of the PDI polypeptide led to secretion of PDI and the C-P4H tetramer in considerable amounts from insect cells, so that it may be concluded that PDI plays a significant role in retaining the C-P4H tetramer within the lumen of the endoplasmic reticulum (Vuori *et al*. 1992c). When the α subunit of a C-P4H is expressed alone in insect cells it forms insoluble aggregates (Vuori *et al*. 1992b). In order to form an active C-P4H the α polypeptide must associate with the PDI polypeptide, and thus another important function of PDI is to keep the α subunit in a catalytically active conformation (Vuori *et al*. 1992b).

2.2.1.2 Catalytic properties

C-P4Hs from all sources require Fe^{2+}, 2-oxoglutarate, O_2 and ascorbate (see Kivirikko & Myllyharju 1998, Kivirikko & Pihlajaniemi 1998, Myllyharju 2003), the K_m values for these cosubstrates being very similar between the animal, viral and plant enzymes (Veijola *et al*. 1994, Helaakoski *et al*. 1995, Annunen *et al*. 1997, 1999, Eriksson *et al*. 1999, Hieta & Myllyharju 2002, Winter *et al*. 2003). The plant and viral P4Hs can use poly(L-proline) as a substrate, whereas this is an efficient competitive inhibitor of some of the animal C-P4Hs. The fact that the K_i value for poly(L-proline) is 1000-fold higher in the human type II enzyme than in the type I enzyme suggests that there are distinct differences between the two in the structures of their peptide-binding sites. The K_m values for the peptide substrates $(Pro\text{-}Pro\text{-}Gly)_{10}$, Gly-Val-Pro-Gly-Val and protocollagen, a protein consisting of non-hydroxylated proα chains of type I procollagen, are 3-6 fold higher in the case of the human type II enzyme (for reviews see Kivirikko & Myllyharju 1998, Kivirikko & Pihlajaniemi 1998, Myllyharju 2003).

The C-P4Hs belong to the group of 2-oxoglutarate-dependent dioxygenases that have a common reaction mechanism which requires the binding of Fe^{2+}, 2-oxoglutarate, O_2 and ascorbate at the catalytic site (see Kivirikko & Myllyharju 1998, Kivirikko & Pihlajaniemi 1998, Myllyharju 2003). In order to identify catalytically important amino acids, homology comparisons were made between several 2-oxoglutarate dioxygenases and a related enzyme, isopenicillin *N* synthase (IPNS). The homology search revealed the presence of two distinct motifs that were 49-71 amino acids apart. Each contained an invariant histidine residue in a conserved position (Myllylä *et al*. 1992). Analysis of IPNS by NMR and other means initially suggested that three histidines are likely to provide the

ligands needed for the binding of the iron atom to a catalytic site (Jiang *et al*. 1991, Ming *et al*. 1991, Randall *et al*. 1993). Site-directed mutagenesis of histidines 412 and 483 in the human α(I) subunit inactivated the type I C-P4H completely, whereas mutation of an additional histidine, His501, inactivated the enzyme by 96% (Lamberg *et al*. 1995, Myllyharju & Kivirikko 1997). Crystallization analysis of the IPNS indicated, however, that instead of a histidine, the third Fe^{2+}-binding ligand is an aspartate in a position +2 with respect to the first Fe^{2+}-binding histidine (Roach *et al*. 1995). Site-directed mutagenesis of Asp414 to alanine or asparagine inactivated the type I C-P4H completely, whereas mutation of Asp414 to glutamate increased the K_m for Fe^{2+} 15-fold and that for 2-oxoglutarate 5-fold (Myllyharju & Kivirikko 1997). Thus there is strong evidence indicating that the amino acids His412, Asp414 and His483 provide the three ligands required for the binding of Fe^{2+} to a catalytic site on the human α(I) subunit.

It has been proposed that 2-oxoglutarate becomes bound to C-P4H via three subsites. Subsite I is a positively charged side chain of the enzyme that ionically binds the C5 carboxyl group of the 2-oxoglutarate, subsite II consists of two *cis*-positioned coordination sites on the enzyme-bound ferrous ion, while subsite III involves a hydrophobic binding site in the C3-C4 region of the cosubstrate (Hanauske-Abel & Günzler 1982, Majamaa *et al.* 1984). Site-directed mutagenesis studies have indicated that the amino acid that ionically binds the C5 carboxyl group of the 2-oxoglutarate in the α(I) subunit is Lys493, since its mutation to alanine inactivated the enzyme completely, whereas the mutation Lys493Arg led to a 15-fold increase in the K_m for 2-oxoglutarate (Myllyharju & Kivirikko 1997). Ascorbate is also believed to become bound to the two *cis*-positioned coordination sites of the enzyme-bound iron (Majamaa *et al*. 1984), but the 2-oxoglutarate-binding site Lys493 does not serve as an additional ligand for ascorbate, as the Lys493Arg mutation did not influence its binding (Myllyharju & Kivirikko 1997).

2.2.1.3 Reaction mechanism

The P4Hs catalyze the formation of 4-hydroxyproline in the presence of the cosubstrates Fe^{2+}, 2-oxoglutarate, O_2 and ascorbate, the 2-oxoglutarate being decarboxylated stoichiometrically during hydroxylation, with one atom of the O_2 incorporated into the succinate and the other into the hydroxy group formed on the proline residue (Figure 1) (Kivirikko & Myllyharju 1998, Kivirikko & Pihlajaniemi 1998, Myllyharju 2003). Kinetic studies have shown that the Fe^{2+}, 2-oxoglutarate, O_2 and peptide substrate become bound to an enzyme molecule in an ordered way, that is the reactants combine with the enzyme and dissociate in an obligatory order and the Fe^{2+} is not released from the enzyme between the cycles (Myllylä *et al*. 1977, Tuderman *et al*. 1977).

Ascorbate is not required in the reaction described above, and the enzyme can catalyze several cycles without it. The hydroxylation eventually stops, however, probably because of formation of the compound $Fe^{3+}{\cdot}O^{-}$, which must be reduced by ascorbate before the next hydroxylation cycle can progress. In the absence of a peptide substrate 2-oxoglutarate is decarboxylated in an uncoupled manner, that is, ascorbate is consumed stoichiometrically and serves as a reducing agent and as an alternative oxygen acceptor (Figure 1) (Kivirikko *et al.* 1992, Kivirikko & Pihlajaniemi 1998, Kivirikko &

Myllyharju 1998). The reaction velocity generated by uncoupled decarboxylation of 2-oxoglutarate is only 1 to 4% of that of a reaction saturated with the peptide substrate (Tuderman *et al.* 1977, Myllylä *et al.* 1984). The uncoupled reactions also take place in the presence of saturating concentrations of the peptide substrates (Figure 1). The biological peptide substrates of C-P4Hs also have regions that do not contain hydroxylatable sequences. When the active site encounters a non-hydroxylatable sequence, uncoupled decarboxylation of 2-oxoglutarate takes place and ascorbate is consumed (Tuderman *et al.* 1977, Myllylä *et al.* 1984, Kivirikko *et al.* 1992).

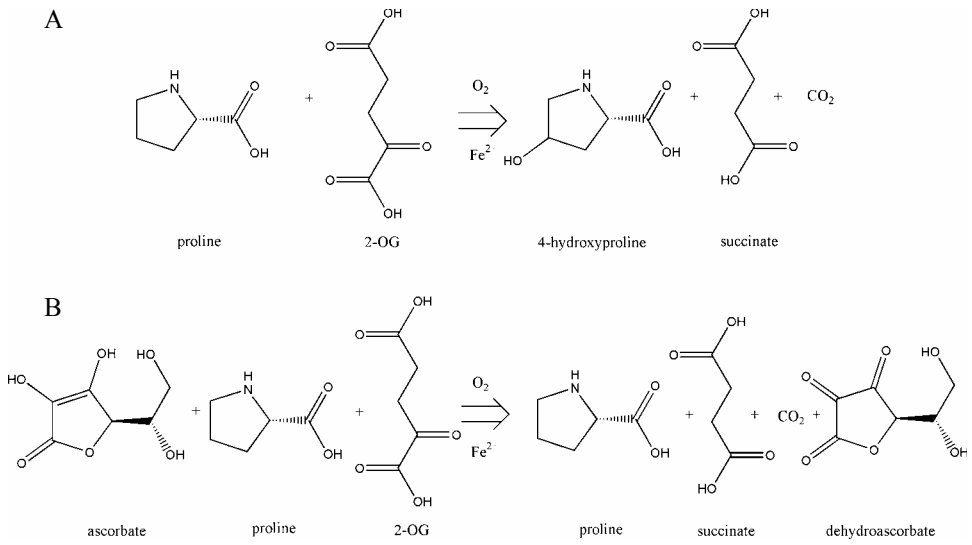


Fig. 1. The reaction catalyzed by the collagen prolyl 4-hydroxylases. 2-Oxoglutarate is decarboxylated stoichiometrically during the hydroxylation of a proline residue (A). In uncoupled reaction cycles ascorbate is consumed stoichiometrically in the presence (B) or absence (not shown) of the peptide substrate.

It is well known that in the reactions catalyzed by the C-P4Hs the K_m values for short peptides such as $(Pro\text{-}Pro\text{-}Gly)_5$ are relatively high, being 1750 μM in the case of the human type I enzyme, while the values for longer peptides are substantially lower, that for $(Pro\text{-}Pro\text{-}Gly)_{20}$ being 50 μM, for example, and that for non-hydroxylated procollagen I chains only 0.2 μM. The decrease can be observed in terms of molar concentrations of both the -X-Pro-Gly- triplets and the peptide (for reviews, see Kivirikko *et al.* 1992, Kivirikko & Pihlajaniemi 1998). It has been proposed that the low K_m values for long peptide substrates are due to the processive mechanism of binding, which requires the presence of two peptide-binding sites and two catalytic sites (de Waal & de Jong 1988, de Jong *et al.* 1991). It has been demonstrated that when one of the two catalytic subunits of the chick C-P4H tetramer is blocked by a photoaffinity analogue of the peptide $(Pro\text{-}Pro\text{-}Gly)_5$, a 5-10 fold increase is found in the K_m for non-hydroxylated procollagen chains, even though the K_m of the photoaffinity-labelled C-P4H for $(Pro\text{-}Pro\text{-}Gly)_5$ is the same as

that for the wild-type enzyme. These data have been interpreted as indicating that both peptide-substrate-binding sites are needed for efficient hydroxylation and low K_m values in the case of long peptide substrates. The processive mechanism of binding theory maintains that upon encountering the C-P4H, the procollagen α chain binds aspecifically to the peptide-binding subsite of one of the two catalytic subunits, after which the other peptide-binding subsite binds to the same α chain, likewise aspecifically. After a series of rapid intersegmental transfers, the enzyme encounters a hydroxylatable X-Pro-Gly segment, which leads to the formation of a specific enzyme-substrate complex. According to this theory the peptide substrate does not dissociate from the enzyme between the hydroxylation cycles but remains bound to at least one of the catalytic subunits. It has been proposed that $(Pro\text{-}Pro\text{-}Gly)_5$ may be too short to become bound to both peptide-binding sites and consequently becomes dissociated from the enzyme between the hydroxylation cycles, which causes the high K_m. These data are in agreement with the observation that the K_m for $(Pro\text{-}Pro\text{-}Gly)_5$ of the C-P4H in which one of the peptide binding sites is blocked is the same as that of the wild-type enzyme (de Waal & de Jong 1988, de Jong *et al*. 1991).

However, contrary to previous suggestions, recent data have shown that the peptide binding site of vertebrate C-P4H tetramers is distinct from the C-terminal catalytic region of the α subunit, the peptide-binding domain of the human type I enzyme being located between the α(I) subunit amino acids Gly138-Ser244 (Myllyharju & Kivirikko 1999). Moreover, surface plasmon resonance studies performed using a recombinant peptide-binding domain Phe144-Ser244 of the human type I enzyme have indicated that longer peptides become bound to individual peptide-binding domains more effectively than shorter ones, the dissociation constants (K_d) determined for various peptides being very similar to their K_m or K_i values for the enzyme tetramers (Hieta *et al*. 2003). These data suggest that the peptide-binding domain is the main determinant of the binding properties of various peptide substrates and inhibitors with respect to the vertebrate C-P4H tetramers.

2.2.2 Nematode collagen prolyl 4-hydroxylases

The collagens in nematodes are involved in two distinct structures, the cuticle and the basement membranes. The basic structure and synthesis of the cuticle is relatively conserved throughout the nematode phylum. The cuticle is an exoskeleton that predominantly consists of collagen-like proteins. The genome of the nematode *C. elegans* contains at least 175 genes, which are predicted to encode cuticular collagen polypeptides and are expressed by specialized epithelial cells, the hypodermis. The cuticular collagen polypeptides typically consist of two collagenous domains, an N-terminal domain containing 8-10 -Gly-X-Y- repeats and a C-terminal domain of 40-42 -Gly-X-Y- repeats. These are flanked and separated by three cysteine-containing non-collagenous domains. The N-terminal non-collagenous region contains a predicted signal sequence cleavage site and a proposed subtilisin cleavage site, which are involved in the formation of a mature polypeptide from a collagen precursor. The expression of collagen genes is cyclically regulated, taking place at high rates during moults and at low rates between the

moult cycles (for reviews, see Johnstone 2000, Page 2001, Myllyharju & Kivirikko 2003).

At least 45 genes have been identified by random mutagenesis, and mutations in these have been shown to affect body shape. Thirteen of the genes encode cuticular collagens (*bli-1*, *bli-2*, *dpy-2*, *dpy-3, dpy-5, dpy-7*, *dpy-8, dpy-10*, *dpy-13*, *lon-3*, *rol-6*, *sqt-1, sqt-3*), two encode basement membrane collagens (*emb-9, let-2*) and 11 encode collagen modifying enzymes (*phy-1, phy-2, phy-3, pdi-2, let-268, bli-4, dpy-11, pdi-3*, *F56C11.1*, *F53G12.3*). Mutations in the cuticular collagen genes yield predominantly three distinct phenotypes: blister (*bli*), dumpy (*dpy*) and roller (*rol*). Nematodes with the blister phenotype are characterized by cuticular swelling and blistering of the cuticular material away from the surface of the worm, whereas dumpy nematodes are short and fat and worms of the roller phenotype rotate around their long axis as they crawl and move in circular paths, in contrast to the sinusoidal wave pattern of movement in wild-type nematodes (for reviews, see Johnstone 1994, 2000, Page 2001, Myllyharju & Kivirikko 2003).

The major component of the basement membranes in *C. elegans* is collagen type IV, a heterotrimer of two α1(IV) chains and one α2(IV) chain. It has been demonstrated that mutations in *emb-9* or *let-2,* which encode the α1(IV) and α2(IV) polypeptides, respectively, result in an embryonically lethal phenotype (Guo *et al*. 1991, Gupta *et al*. 1997).

The synthesized cuticle collagens undergo several co-translational and post-translational modifications. Almost all the proline residues in the Y position of the collagen -Gly-X-Y- sequences are hydroxylated (Kivirikko & Pihlajaniemi 1998). The *C. elegans* genome has two conserved genes encoding polypeptides with a high sequence similarity to the α subunits of vertebrate C-P4Hs, while the products of two other genes show a lower similarity to the vertebrate α subunit sequences (The *C. elegans* Sequencing Consortium 1998). The two conserved genes, *phy-1* and *phy-2,* and one with a low similarity, *phy-3,* have been cloned and characterized, while characterization of the other gene with low homology, *phy-4*, is in progress (Veijola *et al*. 1994, Friedman *et al*. 2000, Hill *et al*. 2000, Winter & Page 2000, Riihimaa *et al*. 2002, Keskiaho K, Kukkola L, Page AP, Winter AD, Kivirikko KI & Myllyharju J, unpublished data). The *C. elegans* PHY-1 and PHY-2 polypeptides consist of 542 and 523 amino acids, respectively, and the signal sequences of 16 additional residues. The PHY-1 and PHY-2 polypeptides show 56.5% identity to each other, the highest conservation being found in the C-terminal region (Friedman *et al*. 2000, Winter & Page 2000). The PHY-1 polypeptide has 44% and 43% identity to human α(I) and α(II) subunits, respectively, and the PHY-2 is 46% identical to the human α(I) and 45% identical to the human α(II) subunit (Winter & Page 2000). All characterized vertebrate C-P4Hs are $\alpha_2\beta_2$ tetramers, but surprisingly, the PHY-1 polypeptide was found to form an active αβ dimer with both the *C. elegans* PDI-2 and human PDI (Veijola *et al*. 1994, 1996). The PHY-1 polypeptide has a C-terminal extension, which may be involved in the formation of the PHY-1/PDI dimer, as its deletion prevents formation of this complex (Veijola *et al*. 1996). The catalytic properties of the PHY-1/PDI dimer are very similar to those of the vertebrate C-P4H tetramers, being closer to those of the vertebrate type II enzyme than to those of the type I enzyme, as the PHY-1/PDI dimer is insensitive to inhibition by poly(L-proline). The K_m values of the PHY-1/PDI dimer for the peptide substrate $(Pro\text{-}Pro\text{-}Gly)_{10}$ and the cosubstrates Fe^{2+},

2-oxoglutarate and ascorbate are similar to those of the vertebrate enzymes (Veijola *et al*. 1994).

The *dpy-18(e499)* strain of *C. elegans* has been shown to be a null mutant of *phy-1*. The worms are viable but phenotypically dumpy, i.e. short and fat (Friedman *et al*. 2000, Hill *et al*. 2000, Winter & Page 2000). The same phenotype was obtained by the double-stranded RNA interference (dsRNAi) method, the resulting phenotype becoming visible in the L4 and adult stages (Hill *et al*. 2000, Winter & Page 2000). Inactivation of the *phy-2* gene by dsRNAi produced no visible morphological effect in wild-type nematodes, whereas *phy-2* dsRNAi performed on the *dpy-18* strain produced an embryonically lethal phenotype, 89% of the embryos failing to hatch. The affected embryos developed normally through gastrulation and elongation, but the abnormal phenotype became evident after cuticle synthesis, whereupon they began to retract to shorter, fatter embryos, which eventually failed to hatch and slowly died (Winter & Page 2000). Disruption of *pdi-2* gene function in wild-type nematodes by dsRNAi resulted in an embryonically lethal phenotype in which 99.9% of the embryos failed to hatch (Winter & Page 2000). These data demonstrated that C-P4H is essential for the normal development of the nematode embryos. The results also gave support to the proposal that the PDI-2 polypeptide represents the only β subunit that associates with the conserved isoforms of the PHY subunits in *C. elegans* (Winter & Page 2000). Using promoter-reporter gene fusion constructs, expression of the *phy-1, phy-2* and *pdi-2* genes was localized to the nuclei of the collagen-synthesizing hypodermal cells in all stages from embryo to adult (Winter & Page 2000).

The third *C. elegans* α subunit homologue characterized, PHY-3, consists of only 295 amino acids, with a proposed signal sequence of 23 additional residues (Riihimaa *et al*. 2002). The amino acid sequence of the processed PHY-3 polypeptide is 17% identical to residues 256-542 in PHY-1 and 18% identical to the corresponding residues in PHY-2 (Riihimaa *et al*. 2002). The PHY-3 polypeptide does not associate with PDI-2, but instead P4H activity was obtained when recombinant PHY-3 was coexpressed with the *C. elegans* PDI-1 (Veijola *et al*. 1996, Riihimaa *et al*. 2002). Due to the small amounts of soluble recombinant PHY-3 available and to the tendency of this polypeptide to aggregate, it has not been possible to determine so far whether PDI-1 forms a dimer or a tetramer with PHY-3, or whether PHY-3 is a monomer, like the *Arabidopsis thaliana* C-P4H (Hieta & Myllyharju 2002), so that PDI-1 is only needed to assist in its folding. Expression of the *phy-3* gene was localized to the spermatheca of stage L4 larvae and adults using promoter-reporter gene fusion constructs and immunofluorescence staining (Riihimaa *et al*. 2002). The spermatheca is a specialized region of the gonad in hermaphrodite nematodes where the oocytes become fertilized (Wood 1988). A homozygous strain with a deletion in the *phy-3* gene had no visible morphological changes in larval-stage worms or adults (Riihimaa *et al*. 2002), but the 4-hydroxyproline content of the early embryo egg shells was markedly reduced. It has therefore been suggested that PHY-3 is involved in the hydroxylation of proline residues in early embryos, possibly of the collagens in their egg shells (Riihimaa *et al*. 2002).

C-P4Hs have also been characterized in the parasitic filarial nematodes *Brugia malayi* and *Onchocerca volvulus* (Merriweather *et al*. 2001, Winter *et al*. 2003). The *B. malayi* PHY-1 polypeptide forms an active, unique C-P4H homotetramer and thus needs no PDI subunit for its solubility or activity (Winter *et al*. 2003). It is expressed in the hypodermal

cells in all developmental stages (Winter *et al*. 2003). At least three, and possibly as many as eight, homologues of the C-P4H *phy* genes are found in the genome of *O. volvulus* (Merriweather *et al*. 2001), and one of these, named Ov-*phy-1,* has been characterized and shown to be expressed at least in stage L3 larvae and adults (Merriweather *et al*. 2001).

Interestingly, all the nematode PHY polypeptides characterized to date are closer homologues to the α(II) subunit of vertebrate type II C-P4H than to the α(I) subunit of the type I enzyme. The nematode C-P4Hs are also relatively insensitive to poly(L-proline) inhibition, which is characteristic of the vertebrate type II enzyme as well (Helaakoski *et al*. 1995, Annunen *et al*. 1997).

2.2.3 Viral and plant prolyl 4-hydroxylases

It was believed for some time that 4-hydroxyproline did not exist in viral proteins and thus a P4H would not be encoded by viral genomes. The genome of a eukaryotic algal virus, *Paramecium bursaria Chlorella* virus-1 (PBCV-1), has nevertheless been shown to encode an active, soluble, monomeric P4H which consists of 242 amino acids, including a putative signal sequence of 32 residues in its N terminus (Eriksson *et al*. 1999). This PBCV-1 P4H shows 20% amino acid sequence identity to the C-terminal half of the human type I C-P4H α(I) subunit, the catalytically critical residues being conserved (Eriksson *et al*. 1999), but no similarity to the peptide-substrate-binding domain which is encoded by amino acids 138-244 in the human α(I) subunit (Myllyharju & Kivirikko 1999). The viral P4H requires Fe^{2+}, 2-oxoglutarate, O_2 and ascorbate as cosubstrates, their K_m values being very similar to those obtained with the vertebrate C-P4Hs (Eriksson *et al*. 1999). It was shown to hydroxylate at least three synthetic peptides corresponding to the proline-rich repeats encoded by the viral genome, including $(Pro\text{-}Ala\text{-}Pro\text{-}Lys)_{2\text{-}10}$, and also poly L-proline and $(Pro\text{-}Pro\text{-}Gly)_{10}$, the latter with a much higher K_m. Interestingly, a decrease in K_m values for peptides with a growing number of repeats of $(Pro\text{-}Ala\text{-}Pro\text{-}Lys)_n$ was observed (Eriksson *et al*. 1999) even though the viral enzyme acts as a monomer (see section 2.2.1.3).

Several plant P4Hs have been identified and characterized to some extent. Early studies on partially purified P4Hs from unicellular (Kaska *et al*. 1987) and multicellular algae (Kaska *et al*. 1988) and the French bean *Phaseolus vulgaris* (Bolwell *et al*. 1985, Wojtaszek *et al.* 1999) have indicated that plant P4Hs exist as monomers. The most distinct difference in catalytic properties between the vertebrate C-P4Hs and plant P4Hs is that the former do not use poly(L-proline) as a substrate, some of them even recognizing it as an effective inhibitor, whereas all the plant P4Hs characterized to date catalyze the hydroxylation of poly(L-proline) effectively (Tanaka *et al*. 1980, Kaska *et al*. 1987, 1988, Kivirikko *et al*. 1992).

Thale cress, *Arabidopsis thaliana*, is the only plant from which a P4H has been cloned and characterized in detail, the genome being shown to contain six open reading frames that encode polypeptides with 21-27% identity to the C-terminal region of the human C-P4H α(I) and α(II) subunits (Hieta & Myllyharju 2002). The cDNA encoding the polypeptide (named At-P4H-1) with highest identity was cloned and expressed as a

recombinant protein. It consists of 283 amino acids, including a predicted non-cleavable signal peptide in its N terminus (Hieta & Myllyharju 2002). The At-P4H-1 resembles the viral PBCV-1 P4H in that it does not show amino acid sequence similarity to the peptide-substrate-binding domain of the vertebrate C-P4Hs (Myllyharju & Kivirikko 1999). The recombinant At-P4H-1 was found to be relatively insoluble, only about 10% of the expressed polypeptide being released into the soluble fraction of the cell homogenate. Its molecular weight proved to be approximately 30 000, lending support to the proposal that plant P4Hs act as monomers (Hieta & Myllyharju 2002). Like the vertebrate C-P4Hs, it requires Fe^{2+}, 2-oxoglutarate, O_2 and ascorbate as cosubstrates, but the K_m value for Fe^{2+} was 4-fold relative to that of the human type I C-P4H, although similar to the K_m values of the P4Hs partly purified from algae (Kaska *et al*. 1987, 1988, Hieta & Myllyharju 2002). The K_m for 2-oxoglutarate was between those of the algal enzymes but 6-fold relative to the human type I C-P4H, and the K_m value for ascorbate was similar to those of the algal P4Hs and human C-P4Hs (Hieta & Myllyharju 2002).

The At-P4H-1 was found to hydroxylate poly(L-proline) and several synthetic peptides based on sequences known to be present in proline-rich cell-wall proteins. Highly surprisingly, it also hydroxylated $(Pro\text{-}Pro\text{-}Gly)_{10}$ effectively, the K_m value being similar to those of the human C-P4Hs (Hieta & Myllyharju 2002). Another interesting characteristic of the recombinant At-P4H-1 was that it hydroxylated the single proline residues present in synthetic peptides representing the human hypoxia-inducible transcription factor HIF-α. These data clearly indicate that the At-P4H-1 does not require a poly(L-proline) type II helix conformation for the hydroxylation of proline residues (Hieta & Myllyharju 2002), as has been previously suggested for plant enzymes (Kivirikko *et al*. 1992).

2.3 HIF prolyl 4-hydroxylases

Oxygen homeostasis is a critical requirement at the systemic level and for a single cell in animals in order to maintain adequate metabolic functions. Hypoxia-inducible factor (HIF) is a transcriptional complex that has a central role in sensing oxygen levels (for reviews, see Semenza 2000, 2001, Fedele *et al*. 2002). This regulation system is evolutionarily conserved, as in addition to humans (Semenza 2001), it has been identified at least in the nematode *C. elegans* (Epstein *et al*. 2001) and the fruit fly *D. melanogaster* (Nambu *et al*. 1996, Bruick & McKnight 2001). The HIF system in simple multicellular animals may have evolved to regulate the cellular energy metabolism in accordance with oxygen availability, while in higher animals the target genes of the HIF system are involved in the regulation of angiogenesis, erythropoiesis, energy metabolism and vasomotor function. The number of known HIF target genes is increasing rapidly (Semenza 2000, 2001).

The HIFs are αβ heterodimers in which both subunits contain basic helix-loop-helix domains that mediate dimerization and DNA binding (Wang *et al*. 1995). Three α subunit isoforms have been identified in humans (see Fedele *et al*. 2002). The abundance of HIF-1α is dependent on the cellular O_2 concentration, whereas HIF-β is expressed constitutively. In normoxia HIF-α is targeted to rapid ubiquitination and proteasomal

degradation, which is regulated by the hydroxylation of Pro564 in HIF-1α, and in some cases also Pro402 (Ivan *et al*. 2001, Jaakkola *et al*. 2001). The resulting 4-hydroxyproline is essential for the binding of HIF-α to the von Hippel-Lindau (VHL) E3 ubiquitin ligase complex and for the subsequent rapid degradation. In hypoxic situations the hydroxylation stops and the HIF-α, which is no longer degraded, becomes bound to HIF-β. The resulting αβ dimer is then transferred to the nucleus, in which it has multiple transcriptional impacts, as mentioned above (Semenza 2000, 2001).

The HIF-1α prolines 564 and 402 are not hydroxylated by the previously known C-P4Hs (Jaakkola *et al*. 2001, Myllyharju 2003), but instead by a novel family of cytoplasmic HIF-P4Hs (Bruick & McKnight 2001, Epstein *et al*. 2001, Ivan *et al*. 2002). The HIF complex in *C. elegans* consists of HIF-1 and AHA-1, which are the *C. elegans* homologues of the HIF-α and β subunits, respectively (Jiang *et al*. 2001). Genetic approaches with mutant worms and database searches to identify candidate enzymes have led to identification of the *C. elegans* HIF-P4H EGL-9, which has been shown to hydroxylate the essential Pro621 in *C. elegans* HIF-α, and three human homologues were subsequently found by using the *egl-9* sequence in database searches (Epstein *et al*. 2001). The human HIF-P4Hs were also independently identified by two other groups (Bruick & McKnight 2001, Ivan *et al*. 2002) and shown to be cytoplasmic and to require O_2, Fe^{2+}, 2-oxoglutarate and ascorbate as cosubstrates, like the C-P4Hs (Bruick & McKnight 2001, Epstein *et al*. 2001, Ivan *et al*. 2002). The HIF-P4H isoenzymes show no overall amino acid sequence similarity to the catalytic α subunits of C-P4Hs, but the three Fe^{2+}-binding residues identified in the C-4PH α subunits, two histidines and an aspartic acid (Lamberg *et al*. 1995, Myllyharju & Kivirikko 1997), are conserved in the HIF-P4Hs, although the lysine that binds the C-5 carboxyl group of 2-oxoglutarate in the C-P4Hs (Myllyharju & Kivirikko 1997) is replaced by an arginine in the HIF-P4H isoenzymes (Bruick & McKnight 2001, Epstein *et al*. 2001, Ivan *et al*. 2002).

In addition to the full-length transcripts, shorter transcripts of the HIF-P4H isoenzymes 2 and 3, encoding inactive polypeptides, have been identified, suggesting novel regulation by alternative splicing (Hirsilä *et al*. 2003). All three isoenzymes were shown to hydroxylate a proline in a 19-residue peptide that corresponded to Pro564 in HIF-1α, but only isoenzymes 1 and 2 hydroxylated a proline corresponding to Pro402, although the K_m was 20-50 times higher (Hirsilä *et al*. 2003). The K_m values of the HIF-P4Hs for O_2 are markedly higher than those of the C-P4Hs, being slightly above its atmospheric concentration, indicating that the HIF-P4H isoenzymes are efficient oxygen sensors (Hirsilä *et al*. 2003).

Genetic defects in the HIF-regulatory system in humans and worms result in severe disorders. Patients suffering from the von Hippel-Lindau syndrome express the HIF target genes constitutively, as they lack any expression of the VHL tumour suppressor protein. This leads to a disorder characterized by blood vessel tumours of the retina and central nervous system, renal carcinoma and pheochromocytoma, for instance (Maxwell *et al*. 1999). *C. elegans* worms lacking the *hif-1* gene are unable to adapt to hypoxia. Where wild-type worms can survive and reproduce in 1% oxygen, most *hif-1* mutants die under these conditions (Jiang *et al*. 2001).

An additional HIF regulating factor is FIH (factor inhibiting HIF), which hydroxylates an asparagine residue in the C-terminal transactivation domain of HIF-α in an oxygen-dependent manner, the hydroxylation serving to block binding of the

transcription coactivator p300/CBP to HIF-α (Mahon *et al*. 2001, Hewitson *et al*. 2002, Lando *et al*. 2002). Like the other known 2-oxoglutarate-dependent dioxygenases, FIH requires Fe^{2+} and 2-oxoglutarate and its function is limited by hypoxia (Hewitson *et al*. 2002, Lando *et al*. 2002).

2.4 Clinical aspects

Hydroxylation of proline residues in collagen chains is essential for collagen stability, and thus also for tissue stability. No mutations have been identified in the C-P4H α subunits in vertebrates, and it was believed until recently that no isoforms of the catalytic α subunit existed, and that absence of the functional C-P4H α subunit would therefore be lethal. Cloning of the second α subunit isoform (Helaakoski *et al*. 1995, Annunen *et al*. 1997) of the vertebrate C-P4H nevertheless raised the possibility that lack of a functional α(I) or α(II) subunit may not lead to lethal phenotypes. Knock-out mouse lines have been generated for both of these genes and analysis of the data is in progress (Holster T, Pakkanen O, Soininen R, Kivirikko KI & Myllyharju J and Pakkanen O, Holster T, Soininen R, Kivirikko KI & Myllyharju J, unpublished data). According to preliminary results, however, α(I) knock-out is embryonically lethal, while α(II) knock-out mice are viable. Knock-out mice will be very useful for analyzing the detailed phenotypic consequences of C-P4H mutations, the abilities of the other isoenzymes to compensate for a lack of one and the specific expression patterns of the isoenzymes at the protein level, and may even lead to the identification of novel diseases caused by mutations in these genes (Myllyharju & Kivirikko 2001, Myllyharju 2003).

Accumulation of fibrous material and scar formation is beneficial and necessary in the wound healing process, for example, but excessive accumulation can cause permanent fibrosis and severely interfere with organ function. Since collagen has a central role in the formation of fibrous material, inhibition of its synthesis could prevent fibrosis. C-P4H also plays a crucial role in the stabilizing of collagen molecules, as non-hydroxylated collagen chains fail to form stable trimers at body temperature. Several studies performed on cultured cells and with *in vivo* models have demonstrated that inhibition of C-P4H prevents the accumulation of collagen. The mechanism of almost all the C-P4H inhibitors developed so far is to bind to the 2-oxoglutarate site, which appears to be very similar between the C-P4H isoenzymes and enzymes from various species, so that the inhibition properties of these compounds for various C-P4Hs are likely to be very similar. By contrast, the peptide-substrate-binding domains show more distant homology between the various C-P4Hs, which might denote higher specificity for inhibitors binding to these domains (see Kivirikko *et al*. 1992, Myllyharju & Kivirikko 2001).

Lymphatic filariasis, also known as elephantiasis, is a disease characterized by dysfunction of the lymphatic system and caused by the parasitic nematodes *Brugia malayi, Brugia timori* and *Wuchereria bancrofti*. More than a billion people are at risk of contracting this disease, and more than 120 million people have already been affected by it. The highest risk exists in India and Africa. It has been estimated that it is the second greatest cause of disability worldwide. Those affected suffer from elephantiasis of the entire leg or arm, genital organs or breast, which may swell up to several times normal

size. At the moment the disease can be treated with two drugs, albendazole and diethylcarbamazine, which have been shown to be effective in killing the adult-stage filarial parasites. The ideal treatment regimens still need to be defined, however (World Health Organization, 2000).

It has been demonstrated that C-P4H function is essential for the normal development of *C. elegans* nematode embryos (Winter & Page 2000), and various inhibitors of vertebrate C-P4Hs have been shown to be toxic to *B. malayi* adults. The inhibitors tested, coumalic acid, phenantrolinone, doxorubicin and daunorubicin, were lethal to the parasitic worms in micromolar concentrations (Merriweather *et al*. 2001). The affected parasitic worms suffered from morphological and biochemical changes consistent with effects on collagen maturation, cuticular repair and maintenance. Worms treated with the C-P4H inhibitor pyridine 2,4-dicarboxylate were normal and viable, however, which is believed to be due to the fact that this compound is a highly polar molecule and thus inefficient at crossing an intact cell membrane. The data suggest that C-P4H inhibitors may exhibit more potent effects on parasitic worms at all developmental stages when they are undergoing a moult (Merriweather *et al*. 2001).

Ischaemia is characteristic of many acute or chronic diseases such as myocardial and cerebral infarction, peripheral arteriosclerosis and diabetes. Stabilization of HIF polypeptides by small molecule inhibitors of HIF-P4Hs would probably have a major impact on factors such as the regulation of angiogenesis and erythropoiesis, and it is thus believed that the inhibitors could be beneficial in the treatment of these diseases (Bruick & McKnight 2001, Epstein *et al*. 2001, Ivan *et al*. 2002). A number of competitive inhibitors of C-P4Hs with respect to 2-oxoglutarate also inhibit HIF-P4Hs (Hirsilä *et al*. 2003), but considerable differences were observed in K_i values between human type I C-P4H and the HIF-P4Hs, and even between the three human HIF-P4H isoenzymes (Hirsilä *et al*. 2003). Thus it should be possible to develop inhibitor molecules that are highly specific for HIF-P4Hs and even for individual HIF-P4H isoforms.

2.5 The nematode *C. elegans* as a model organism

The nematode *C. elegans* was selected as an experimental model organism by Sydney Brenner in the early 1960's in order to explore animal development and the underlying principles of the functioning of the nervous system in a simple metazoan. *C. elegans* is a free-living, non-parasitic nematode with a life-cycle of 3.5 days at 20°C. A single adult hermaphrodite gives birth to at least 300 offspring by self-fertilization. The adult worm is 1 mm in length and about 80 μm in width and can easily grow in Petri dishes on agar medium when fed on a diet of *Escherichia coli*. The number of cells in an adult hermaphrodite is invariant, each worm consisting of 959 somatic cells, of which 302 are neurons. A precise, detailed map of the *C. elegans* genome is available, and the genome has been fully sequenced. Research on the nematode *C. elegans* was granted the Nobel Prize in 2002 (for reviews, see Culetto & Sattelle 2000, Hariharan & Haber 2003).

Humans, and vertebrates in general, seem to be very far removed from the worms in many aspects. But what do we have in common? The worms have many cell types that are similar to ours: neurons, muscle cells, gut and excretory cells, but above all we share

many conserved genes and cellular mechanisms. The on-going accumulation of data on *C. elegans* is beginning to demonstrate that most of the important biological processes have remained essentially unchanged during evolution. Worms suffer from many diseases that are similar to human ones, including cancer, neurodegeneration, infectious diseases and disorders of physiological control, and ageing, but their organic physiology differs from ours in many respects (for reviews, see Culetto & Sattelle 2000, O'Kane 2003).

Many human diseases are now known to originate from a mutation in a single gene. The distribution of a gene product in tissues can be studied in many ways, using mRNA analysis, immunostaining and tissue samples as materials for its isolation, and the effect of the absence of a gene product can be explored by generating knock-out lines of mice or rats. Bioinformatics can provide clues to the biological role of a gene product and its possible association with diseases by predicting functional domains in it. These approaches do not, however, provide data on the role of a given gene product in the functional context of the signalling pathways that it is involved in (Culetto & Sattelle 2000).

What benefits, then, can a worm provide when modelling human diseases? Biological processes are easier to unravel in small, transparent animals than in humans, especially when procedures are required that are not applicable to higher animals or cannot be used for ethical reasons. The short generation times and the ability of worms to produce progenies in astonishing amounts are undoubtedly distinct benefits, too. Mutant strains can be generated efficiently and the resulting phenotypes can be identified rapidly. Strains that overexpress the gene to be studied or lack the gene function can also be generated and analysed relatively quickly (Hariharan & Haber 2003).

It has been estimated that about 42% of the human disease genes have an orthologue in the genome of *C. elegans* including those associated with Alzheimer disease, hereditary non-polyposis colon cancer, spinal muscular atrophy, juvenile Parkinson's disease and many, many others. There are several approaches by which a worm model can be used to study human diseases, e.g. the human gene can be overexpressed in certain cells of *C. elegans* by means of muscle or neuron-specific promoters and the expression of genes at the RNA level in the resulting worm population can be studied by microarray analysis, which can reveal a batch of up and down-regulated genes. In simplified terms, those up or down-regulated genes that have a human homologue can provide a baseline for further studies aimed at identifying members of the signalling cascades activated in the disease and candidates for drug therapy. This approach can be useful in the case of a neurological disease such as Alzheimer disease, in which it is impossible to study the pathological early stages because brain samples are available for analysis only after death (Link *et al*. 2003, O'Kane 2003).

Another approach to the study of human diseases is to inactivate the function of a *C. elegans* gene orthologue by double-stranded RNA interference (dsRNAi). This method was first introduced in worms and is nowadays an effective tool that can be used for silencing the expression of a gene of interest in other species as well. Double-stranded RNA for the target gene is introduced into the worms either by microinjection or ingestion of bacterially expressed dsRNAs or by soaking them in a liquid containing the target dsRNA (Fire *et al.* 1998, Tabara *et al*. 1998, Timmons *et al*. 2001), which can activate mechanisms in the nematode cells that target the degradation of cognate cytoplasmic mRNAs and thus effectively evoke gene silencing at the post-transcriptional

level (Montgomery *et al*. 1998). The dsRNAi method is a potent means of searching for "modifier genes", which enhance or suppress the mutant phenotype. Worms with a certain mutant phenotype can be maintained in thousands of small wells, each containing bacteria expressing a different dsRNA. The phenotypes that are restored or enhanced are indicated for further study because the dsRNA interacting with them probably belongs to the same signalling cascade as the mutant gene which generated the deviant phenotype (for a review, see Hariharan & Haber 2003).

3 Outlines of the present research

The collagen prolyl 4-hydroxylases form an enzyme family with several isoenzymes in a number of species, including the human, rat, mouse, nematodes and fruit fly. This work, reports on the cloning and characterization of a third human, mouse and rat C-P4H α subunit isoform, providing data that should establish a basis for future attempts to unravel the roles of the various vertebrate C-P4H isoenzymes *in vivo,* e.g. by means of knock-out mouse technology.

In the second part of the work certain novel *C. elegans* C-P4H forms, a PHY-1/PHY-2/$(PDI\text{-}2)_2$ mixed tetramer and PHY/PDI dimers, are characterized and their functions are found to be critical for the normal growth and development of the nematode. Furthermore, the *C. elegans* mixed tetramer is used as a model to study the roles of the two catalytic α subunits in a C-P4H tetramer. Novel data on the function of these sites are reported and some of the previously suggested hypotheses regarding the reaction mechanism are ruled out.

The specific goals were:

1. to clone a third human, rat and mouse C-P4H α subunit isoform, to study the tissue distribution of human α(III) mRNA and to express recombinant human type III C-P4H in order to determine its catalytic properties,
2. to study the assembly of the *C. elegans* C-P4H subunits PHY-1, PHY-2 and PDI-2 *in vitro* and *in vivo*, to determine the catalytic properties of the novel C-P4H forms and to identify critical regions in the PHY polypeptides for enzyme assembly, and
3. to study the roles of the two catalytic sites in the C-P4H tetramer.

4 Materials and methods

Detailed descriptions of the materials and methods are presented in the original articles I-III.

4.1 Isolation of cDNA clones, and generation of recombinant baculoviruses and stably transfected mammalian cell lines (I-III)

To isolate cDNA clones for the human C-P4H α(III) subunit, the human umbilical vein endothelial cell (HUVEC) lambda cDNA library (Stratagene) was screened using a probe which was a 162-bp PCR product. The probe was amplified from a Human Foetal Marathon-Ready cDNA pool (BD Biosciences) based on an expressed sequence tag (EST) clone showing similarity to the human α(I) and α(II) subunits. 13 positive clones were obtained and the sequences were analysed. The 5' end of α(III) cDNA was obtained by 5' RACE of HUVEC cDNA, which was generated using the SMART RACE cDNA amplification kit (BD Biosciences), and total RNA isolated from cultured HUVEC cells with the RNeasy Midi kit (Qiagen). The full-length human α(III) cDNA was obtained by first cloning a fragment extending from bp 184 to the polyA tail into pUC18 using the SureClone ligation kit (Amersham Biosciences). This construct was then digested with *Bam*HI-*Apa*I and a similarly digested α(III) fragment covering bp 1 to 418 was ligated to it. The rat and mouse α(III) cDNAs were assembled by consecutive overlapping 5' and 3' RACE reactions using various rat and mouse Marathon-Ready cDNA pools (BD Biosciences) as templates and initial 5' and 3' RACE primers based on a 169 bp rat EST sequence.

The full-length human α(III) cDNA was cloned into the baculovirus vector pVL1393 (Pharmingen) and α(III) cDNA lacking sequences coding for the signal peptide was cloned into pACGP67A (Pharmingen) in frame with the baculovirus GP67 signal sequence.

Mammalian expression vectors were generated by cloning the full-length human PDI cDNA into the pCDNA3.1(+) vector (Invitrogen) and the full-length human α(III) and

α(I) cDNAs into pCDNA3.1(+)Hygro vectors (Invitrogen). The pCDNA3.1(+)PDI construct was transfected into human embryonic kidney HEK-293 cells using PolyFect (Qiagen), and stably transfected cells were selected with Geneticin (Gibco BRL). The pCDNA3.1(+)Hygro vectors encoding the α(III) and α(I) polypeptides were transfected into the stable HEK-293 cell line expressing recombinant human PDI, and stable cell lines were selected with hygromycin (Gibco BRL). A control cell line was established by stable transfection of pCDNA3.1(+)HygroLacZ (Invitrogen).

The full-length *phy-2* cDNA was cloned into a pVL1392 expression vector (Invitrogen). Expression constructs for hybrid PHY polypeptides were obtained by preparing two PCR fragments using pVL1392-*phy-1* (Veijola *et al.* 1994) and pVL1392-*phy-2* as templates and ligated into pVL1392. Hybrid A consisted of the PHY-1 signal peptide, PHY-1 amino acids 1–120 and PHY-2 amino acids 123–523, hybrid B of the PHY-1 signal peptide, PHY-1 amino acids 1–271 and PHY-2 amino acids 268–523, hybrid C of the PHY-2 signal peptide, PHY-2 amino acids 1–122 and PHY-1 amino acids 121–543, and hybrid D of the PHY-2 signal peptide, PHY-2 amino acids 1–267 and PHY-1 amino acids 272–543 (figure 7 in II). The single amino acid mutagenesis reactions of the *C. elegans phy-1* (Veijola *et al*. 1994) and *phy-2* cDNAs were carried out in a pVL1392 vector (Invitrogen) using the QuikChange™ XL Site-Directed Mutagenesis Kit (Stratagene). The PHY-1 Asp407 and Lys486 residues were converted individually to asparagine and alanine, respectively, and likewise the PHY-2 Asp405 and Lys484 residues.

All the baculovirus constructs were cotransfected into *Spodoptera frugiperda* Sf9 insect cells with a modified *Autographa californica* nuclear polyhedrosis virus DNA (BaculoGold, PharMingen) and the resultant virus pools were collected and amplified three times (Crossen & Gruenwald 1998).

4.2 Expression and analysis of the recombinant proteins in insect and mammalian cells (I-III)

In all insect cell studies (I-III) the cells (Sf9 or High Five, Invitrogen) were cultured as monolayers in TNM-FH medium (Sigma) supplemented with 10% insect cell qualified foetal bovine serum (Invitrogen) at 27°C. To produce recombinant proteins, cells seeded at a density of $5x10^6$ per 100-mm plate were infected with the recombinant viruses at a multiplicity of 5, harvested 72 h after infection, homogenized in a buffer containing Triton X-100 and centrifuged. Samples of the resulting supernatants and the remaining pellets solubilized in 1% SDS were analysed by denaturing SDS-PAGE or non-denaturing PAGE followed by Coomassie Blue staining or Western blotting.

The stably transfected mammalian cell lines were cultured in Dulbecco's modified Eagle's medium (Biochrom) supplemented with 10% foetal calf serum (BioClear) and 50 μg/ml of ascorbic acid at 37ºC. The cells were harvested at confluency, homogenized in a buffer containing Triton X-100 and analysed as above.

To detect human PDI, α(I) or α(III) in the Western blots, the blots were analysed with the monoclonal PDI antibody 5B5 (Dako), the polyclonal α(I) antibody K17 (Veijola *et*

al. 1996) or the monoclonal antibody VTT1081 against the α(III) subunit generated by immunizing mice with a recombinant α(III) polypeptide (Technical Research Centre of Finland). To detect the PHY-1, PHY-2 and PDI-2 polypeptides in the Western blots and in the immunolocalization studies (see below), three polyclonal antibodies were generated in rabbits against synthetic peptides representing the C-terminal residues of PHY-1, PHY-2 and PDI-2 (Sigma and Genosys). The cuticle collagen-specific DPY-7 monoclonal antibody was a gift from Dr. Iain Johnstone, University of Glasgow.

4.3 Immunoprecipitation, N glycosidase F treatment and gel filtration studies (I)

Native immunoprecipitation was performed to analyse whether the α(III) and PDI polypeptides are present in the same molecule. Protein G Sepharose 4 Fast Flow was used according to the instructions provided by the manufacturer (Amersham Biosciences). The cell lysates were pre-cleared and immunoprecipitated with the antibody VTT1081 against the α(III) subunit or anti-flag (Sigma) as a negative control. The sepharose and bound antibody-protein complexes were washed three times with a Triton X-100 buffer (see section 4.2), after which the precipitate was analysed by SDS-PAGE under reducing conditions, followed by Western blotting using the α(III) and PDI antibodies.

N glycosidase F treatment was carried out in the soluble fraction of homogenized cells coexpressing the recombinant human α(III) and PDI or α(I) and PDI polypeptides according to the instructions provided by the manufacturer (Roche).

Gel filtration was performed in a calibrated Superdex 200 column (Amersham Biosciences) using the soluble fraction of homogenized cells coexpressing recombinant human α(III) and PDI or α(I) and PDI polypeptides as a sample.

4.4 Analysis of the expression of the human α(III) subunit mRNA (I)

The Northern blots Human MTN Blot (BD Clontech) and Real^{TM} Human Fetal mRNA Blots I and II (Invitrogen) were hybridized under the conditions specified by the manufacturer with a 2250-bp α(III) or 1440-bp α(I) cDNA fragment as the probe, the blots being exposed for 72 and 4 h, respectively.

Total RNA from human foetal epiphyseal cartilage and human fibroblasts (N-09) was isolated using the RNeasy Midi kit (Qiagen), RT-PCR was performed using a SMART RACE cDNA Amplification Kit (BD Clontech), and PCR analysis of the cartilage and fibroblast cDNAs, Human MTC Panel I and Human Fetal MTC Panel (BD Clontech) was performed according to the manufacturer's protocol using primer pairs specific to the human α(I) and α(III) sequences. In the case of the cartilage and fibroblast cDNAs α(II) sequence specific primers were also used. 5-μl aliquots were taken after 26, 30, 34, 38 and 42 cycles and analysed on 1% agarose gels.

To search for possible alternatively spliced exons in human C-P4H α(III) cDNA, the α(III) cDNA was amplified from various human MTC panel samples and Marathon-Ready cDNAs (BD Biosciences) as templates using two primer pairs that span all the exon-intron boundaries of the cDNA.

4.5 Detection and characterization of C-P4Hs in *C. elegans* lysates and immunolocalization of *C. elegans* C-P4H (II)

Wild-type (N2), *dpy-18*(*e364*) and *phy-2*(*ok177*) strains were received from the *C. elegans* Genetics Center. The deletion strain *phy-3*(*ok199*) was generated by the *C. elegans* Genome Deletion Consortium (Riihimaa *et al*. 2002).

Native worm extracts were prepared by washing the mixed stage nematodes several times with a protease inhibitor mixture in ice-cold phosphate-buffered saline, centrifuging them and freezing them at -70°C. The pellets were resuspended in a Triton X-100 buffer on ice, homogenized and centrifuged, and the supernatants were analysed by non-denaturing PAGE followed by Western blotting with antibodies to PHY-1, PHY-2, and PDI-2.

To study the localization of the PHY-1, PHY-2 and PDI-2 polypeptides in the nematode, wild-type embryos were washed extensively, pipetted onto poly-L-lysine-coated slides and permeabilized by freeze-cracking. Samples were then blocked and probed with combinations of monoclonal anti-DPY-7 and polyclonal PHY-1, PHY-2, and PDI-2 antibodies and then washed extensively and incubated in a mixture of Alexa Fluor 594 goat anti-rabbit IgG and Alexa Fluor 488 goat anti-mouse IgG (Molecular Probes). After washing, they were viewed under epifluorescence on a Zeiss Axioskop 2 microscope.

4.6 2-Oxoglutarate analogue inhibition assays in nematodes *in vivo* (II)

To study 2-oxoglutarate analogue inhibition in the nematodes, small Petri plates were prepared with NGM-agarose without salts and seeded with OP50 bacteria. They were then inoculated with pyridine 2,4-dicarboxylate or pyridine 2,5-dicarboxylate to final concentrations ranging from 60 μM to 10 mM. Fifteen 2-fold embryos of each strain, *dpy-18*(*e364*), *phy-2*(*ok177*) and *phy-3*(*ok199*), were added to each plate, incubated at 20 or 25 °C, and viewed twice daily over several days.

4.7 RNAi experiments on *C. elegans* (II)

Additional effects of *phy-3* inactivation were studied by means of RNAi experiments performed on wild-type *C.elegans* and the deletion strains *dpy-18(e364)* and *phy-2(ok177)*. Double-stranded RNA of the *phy-3* coding sequence was produced *in vitro* (Winter & Page 2000) and introduced into the nematodes by two standard mechanisms, microinjection and bacterially mediated RNAi (Fire *et al*. 1998, Timmons *et al*. 2001), and the resulting progeny were analysed.

4.8 P4H activity assays (I-III)

The P4H activity of the Triton X-100 soluble proteins expressed in mammalian or insect cells and of the nematode lysates was assayed by a method based on the formation of 4-hydroxy[^{14}C]proline in a [^{14}C]proline-labelled substrate consisting of non-hydroxylated procollagen polypeptide chains, or by a method based on the hydroxylation-coupled decarboxylation of 2-oxo[1-^{14}C]glutarate (Kivirikko & Myllylä 1982). The peptide substrates used in the latter method were $(Pro\text{-}Pro\text{-}Gly)_5$, $(Pro\text{-}Pro\text{-}Gly)_{10}$ (Peptide Institute) or recombinant non-hydroxylated 100 or 510-amino acid fragments of the α1 chain of type I collagen (see below).

The latter chain fragments were expressed in *Pichia pastoris* by cloning cDNA fragments encoding polypeptides starting from the proα1(I) chain amino acids 683 and 1093, respectively, and extending to the last amino acid of the collagenous domain into the pPIC9K expression vector (Invitrogen) in frame with the yeast α-mating factor secretory signal. The constructs were linearized and electroporated into a GS115 *Pichia* strain (Invitrogen) according to the manufacturer's instructions. The recombinant strains were cultured in 100-ml shaker flasks in a buffered glycerol complex medium and expression was induced in a buffered minimal methanol medium, methanol being added every 12 h to a final concentration of 0.5%. The culture medium was collected 60 h after induction and concentrated. The recombinant type I collagen polypeptide fragments were analyzed by denaturing SDS-PAGE followed by Coomassie Blue staining.

5 Results

5.1 Molecular cloning and characterization of the human C-P4H α(III) subunit and expression of an active $[\alpha(III)_2]\beta_2$ tetramer in mammalian cells (I)

5.1.1 Cloning of the human, rat and mouse α(III) polypeptides and analysis of the amino acid sequences

A sequence homology search identified a number of human ESTs representing a gene product with similarity to the conserved catalytic regions of the human C-P4H α(I) and α(II) subunits. The EST AA116081 sequence was used to design primers for 5' and 3' RACE reactions and to amplify a probe for the screening of cDNA libraries. The resulting α(III) cDNA clones cover 43 bp of the 5' untranslated sequence, a 1635-bp open reading frame and 591 bp of the 3' untranslated sequence, including the polyadenylation signal. Full-length rat and mouse α(III) cDNAs were obtained by RACE reactions using primers based on a 169-bp rat sequence showing similarity to the human α(III) cDNA.

The processed human and rat α(III) subunits were 525 amino acids in length and the mouse α(III) 520 amino acids. The human and rat α(III) cDNAs encode additional signal peptides of 19 residues, while the length of the mouse signal peptide is 22 residues. The three α(III) polypeptides are 91-95% identical. The overall amino acid sequence identity between the processed human α(III) and α(I) subunits is 35%, and that between α(III) and α(II) 37% (figure 1 in I). The identity is not distributed equally, however, but is highest within the catalytic C-terminal region. All four catalytically critical residues (Myllyharju & Kivirikko 1997) are conserved in the three human α subunits (figure 1 in I). The identity between the human α(I) peptide-substrate-binding domain (Myllyharju & Kivirikko 1999) and the corresponding amino acids of the α(II) and α(III) subunits is 57% and 35%, respectively, and that between the α(II) and α(III) is 34%. Five conserved cysteine residues are found in the human α(I), α(II) and α(III) subunits, the α(II) and

α(III) subunits having an additional cysteine between the conserved cysteines 4 and 5, and 1 and 2, respectively (figure 1 in I).

5.1.2 Expression of the α(III) mRNA in various tissues

The highest α(III) mRNA expression levels in the Northern blot analyses were seen in the placenta, adult liver and foetal skin, low levels being found in the foetal liver, lung and muscle (figure 2A in I). PCR analysis of cDNA samples from various human tissues demonstrated that highest α(III) mRNA expression levels were in the placenta and foetal kidney, liver and lung (figure 2B in I). The expression levels of the α(III) mRNA were much lower than those of the α(I) mRNA in all the tissues studied.

5.1.3 Characterization of the human α(III) gene

The exon-intron organization of the human α(III) gene is very similar to those of the α(I) and α(II) genes, with just a few changes in the exon-intron boundaries and the lengths of the exons (figure 3A in I). The possible existence of alternatively spliced exons was studied by careful sequence homology analysis of all the introns in the human α(III) GenBank clone AC006595 and PCR analysis of various tissues using two primer pairs to generate PCR products spanning all the exon-intron boundaries. Unlike the α(I) and α(II) mRNAs, the α(III) mRNA seems not to be subject to alternative splicing, at least not in any of the tissues studied.

5.1.4 Expression and characterization of recombinant human type III C-P4H

A recombinant α(III) polypeptide containing either its own signal peptide or the baculoviral signal peptide GP67 was coexpressed with PDI in insect cells, but no P4H activity was generated and no association was detected between these two polypeptides. N-terminal sequencing of the recombinant α(III) polypeptides indicated that the native signal peptide was not cleaved, while the GP67 signal peptide was cleaved incorrectly.

Stable transfection of mammalian HEK-293 with recombinant α(III) and PDI or α(I) and PDI polypeptides was found to lead to 3.5 and 6-fold increases in the P4H activity, respectively, relative to a control transfection (table I in I). Immunoprecipitation with the α(III) antibody was found to precipitate both the α(III) and PDI polypeptides from the soluble fraction from cells coexpressing these two polypeptides, while the α(I) and PDI polypeptides were not precipitated from cells expressing these two (figure 5 in I), indicating that the α(III) and PDI polypeptides are very likely to be present in the same

molecule. Further studies by gel filtration showed that the elution position of the P4H activity was identical for cell lines coexpressing either the α(III) and PDI or the α(I) and PDI polypeptides, and also identical to that of purified recombinant human type I C-P4H (figure 6A in I). The gel filtration fractions showing the highest activity levels were analysed by SDS-PAGE and non-denaturing PAGE followed by Western blotting. Non-denaturing PAGE of the fractions from cells coexpressing α(III) and PDI showed a band that had a mobility similar to that of the purified recombinant type I C-P4H tetramer and could be stained with the α(III) and PDI antibodies, but not with the α(I) antibody (figure 6B in I). It thus seems very likely that the type III C-P4H has a similar subunit composition to the type I and type II C-P4Hs and is an $[\alpha(III)]_2\beta_2$ tetramer in which PDI serves as the β subunit.

To study the possibility of the non-conserved cysteine 226 of the α(III) subunit participating in the formation of an interchain disulphide bond, soluble fractions from cells coexpressing the α(III) and PDI polypeptides were analyzed by SDS-PAGE under non-reducing and reducing conditions. The α(III) polypeptide was found to migrate in the position of a polypeptide monomer under both conditions (figure 7A in I), and thus it cannot be involved in the formation of interchain disulphide bonds. N glycosidase F treatment showed that the recombinant human α(III) polypeptide is glycosylated in the HEK293 cells (figure 7B in I). No difference in enzyme activity was found between the N glycosidase F-treated and non-treated samples, indicating that glycosylation of the α(III) polypeptide does not play a role in the catalytic activity of type III C-P4H.

5.1.5 Catalytic properties of type III C-P4H

The K_m values of the type III C-P4H for the cosubstrates Fe^{2+}, 2-oxoglutarate and ascorbate were found to be essentially identical to those of the type I and type II C-P4Hs (Table II in I). The IC50 for pyridine-2,4-dicarboxylate was also very similar to those of the type I and type II C-P4Hs, but that for pyridine-2,5-dicarboxylate was 4-7-fold (Table II in I). The K_m of the type III C-P4H for the substrate $(Pro\text{-}Pro\text{-}Gly)_{10}$ was slightly higher than that of the type I C-P4H, but distinctly lower than that of the type II C-P4H (Table II in I). The IC50 of the type III C-P4H for poly(L-proline), M_r 5000, was 5 times that of the type I C-P4H but only one tenth of that of the type II C-P4H (table II in I).

5.2 Characterization of *C. elegans* C-P4H forms assembled from the PHY-1, PHY-2 and PDI-2 polypeptides (II)

5.2.1 Expression of recombinant C. elegans PHY-1, PHY-2 and PDI-2 polypeptides and characterization of the assembled C-P4Hs

The recombinant *C. elegans* PHY-2 polypeptide was expressed in insect cells and Triton X-100 and SDS-soluble proteins from the cell homogenates were analysed by SDS-PAGE under reducing conditions. Efficient extraction of the recombinant PHY-2 from the cell homogenates required the use of 1% SDS (figure 1A in II). To study whether the *C. elegans* PHY-2 assembles with the *C. elegans* PDI-1, PDI-2, or human PDI polypeptides, insect cells were coinfected with viruses coding for recombinant PHY-2 and one of the PDI subunits. Analysis by non-denaturing PAGE demonstrated that PHY-2 does not become associated with any of the PDI subunits (figure 1B-C in II), while PHY-1 was shown to form an αβ dimer with either the *C. elegans* PDI-2 or human PDI, as reported previously (Veijola *et al*. 1994, 1996).

To study whether PHY-2 can assemble into mixed tetramers with PHY-1 and PDI-1, PDI-2, or human PDI, both types of PHY polypeptide were coexpressed in insect cells together with one of the PDI subunits. Coexpression of PHY-2 and PHY-1 with PDI-2 led to the formation of an active P4H (Table I in II) with a mobility similar to that of the human type I C-P4H tetramer (figure 2A in II). Immunostaining with polyclonal antibodies against PHY-2 and PHY-1 showed that both polypeptides were present in this tetramer (figure 2B and C in II). The data obtained show that a unique mixed PHY-1/PHY-2/$(PDI\text{-}2)_2$ C-P4H tetramer is assembled from the *C. elegans* polypeptides. In addition, a faint band corresponding to a PHY-1/PDI-2 dimer was observed (figure 2C in II), but no PHY-2/PDI-2 dimer was detected (figure 2B in II).

To study whether PHY-1 and PHY-2 can form a mixed tetramer with P4H subunits from other species, the PHY-1 or PHY-2 were coexpressed in insect cells with viruses coding for human C-P4H α(I) or α(II) or the *Drosophila* α subunit in combination with *C. elegans* PDI-1, PDI-2, human PDI, or *Drosophila* PDI. Non-denaturing PAGE analysis demonstrated that assembly of a mixed enzyme tetramer is species-specific, as a band corresponding to an enzyme tetramer was seen only when PHY-2 was coexpressed with PHY-1 and *C. elegans* PDI-2 (figure 6 in II).

The catalytic properties of the recombinant *C. elegans* mixed P4H tetramer (table II in II) were very similar to those of the dimer formed by PHY-1 and human PDI (Veijola *et al*. 1994) and the human type I C-P4H (Myllyharju & Kivirikko 1997). The only difference was that the K_m value of the mixed tetramer for 2-oxoglutarate was slightly higher, about 80μM (table II in II). The mixed tetramer resembles human type II C-P4H (Helaakoski *et al*. 1995, Annunen *et al*. 1997) and the *C. elegans* PHY-1/human PDI dimer (Veijola *et al*. 1994) in that it is not efficiently inhibited by poly(L-proline).

5.2.2 Analysis of C. elegans C-P4Hs in vivo

The insect cell coexpression studies showed that the mixed tetramer is the major form of recombinant *C. elegans* C-P4H, assembly of the PHY-1/PDI-2 dimer being much less efficient. To study whether the mixed tetramer is also the major form of *C. elegans* C-P4H *in vivo*, native extracts were prepared from wild-type and *phy* null-mutant nematodes and analyzed by non-denaturing PAGE followed by immunostaining. The mixed tetramer was found to be the major C-P4H form in the wild-type nematodes and only a small amount of the PHY-1/PDI-2 dimer was detected (figure 3 in II). A new PHY-2 immunoreactive band was detected in the *phy-1* null mutant strain *dpy-18(e364)*, however, indicating the assembly of a novel PHY-2/PDI-2 dimer (figure 3 in II). In the *phy-2* null mutant strain a band corresponding to a PHY-1/PDI-2 dimer was detected, as expected, but assembly of this dimer was markedly increased by comparison with the wild-type strain (figure 3 in II).

The P4H activities of the *phy-1* null and *phy-2* null nematodes were found to be reduced to 0.7%-2.5% and 54-57% of that of the wild-type nematodes, respectively.

The expression patterns of the *phy-1*, *phy-2*, and *pdi-2* genes at the levels of the encoded polypeptides were studied by immunostaining of wild-type embryos with subunit-specific antibodies. The collagenous cuticle has not yet formed in the pre-elongated 1.5-fold embryos, and identical immunostaining patterns in the endoplasmic reticulum of hypodermal cells were obtained with the C-P4H subunit antibodies and the cuticle collagen-specific antibody anti-DPY-7 (figure 4A in II). In the elongated embryos the first larval cuticle had been synthesized, and at this stage the hypodermal ER location was maintained for all three P4H subunits, whereas the cuticle collagen DPY-7 had already been secreted and was fully incorporated into the developing cuticles (figure 4B in II).

The *in vivo* effects of the 2-oxoglutarate analogues pyridine 2,4-dicarboxylate and pyridine 2,5-dicarboxylate were examined in a plate assay performed on live wild-type and *phy*-null nematodes, including a *phy-3* null strain. Slight effects were noted in the wild-type, *phy-2* null, and *phy-3* null nematodes, including mild moulting defects at 5-10 mM concentrations of the compounds. Severe effects were noted in the *phy-1* null strain for both compounds at concentrations of 0.5-4 mM, however, such as embryonic lethality, sterility, severe dumpiness and moulting defects (figure 5 in II).

5.2.3 Identification of critical regions in the PHY-1 and PHY-2 polypeptides for the assembly of the mixed C. elegans C-P4H tetramer

To identify regions in the PHY-1 and PHY-2 polypeptides that are required for mixed tetramer assembly, four baculoviruses coding for the hybrid PHY polypeptides A-D (figure 7 in II) were coexpressed with the PHY-1 or PHY-2 and PDI-2 polypeptides in insect cells or with PDI-2 in various combinations. The formation of enzyme tetramers was analysed by non-denaturing PAGE followed by Western blotting with the PHY-1 and

PHY-2 antibodies (figures 7 and 8 in II). The results obtained indicated that the PHY-2 residues Asp1-Leu122 are sufficient for a hybrid polypeptide to act as a PHY-2-like subunit in mixed tetramer assembly, whereas the critical region for a hybrid polypeptide to act as a PHY-1-like unit in mixed tetramer assembly is located between the PHY-1 residues Gln121 and Ala271 (figure 7 in II).

5.3 Characterization of the effect of peptide substrate chain length on the kinetic properties of a C-P4H tetramer and dimer and the effect of inactivation of one of the catalytic sites in a C-P4H tetramer (III)

5.3.1 Effect of increasing peptide chain length on the K_m values of a C-P4H tetramer and dimer

The K_m values of C-P4Hs for peptide substrates decrease distinctly with increasing peptide chain length (Kivirikko *et al*. 1992, Kivirikko & Pihlajaniemi 1998). According to the processive mechanism of binding theory, an enzyme with two peptide-binding sites on one molecule should be much more efficient in hydroxylating long peptide substrates than an enzyme with only one such site, and furthermore, this effect should not be evident in the hydroxylation of short peptides. To study the suggested co-operation of the two peptide-binding sites in more detail, a recombinant *C. elegans* PHY-1/PHY-2/$(PDI\text{-}2)_2$ tetramer, i.e. a P4H with two peptide-binding and catalytic sites, and the PHY-1/human PDI dimer, i.e. a P4H with one peptide-binding and one catalytic site, were produced in insect cells (figure 1 in III) and their K_m values for peptide substrates of four lengths were determined. The peptide substrates used were the synthetic peptides $(Pro\text{-}Pro\text{-}Gly)_5$, $(Pro\text{-}Pro\text{-}Gly)_{10}$, and recombinant non-hydroxylated 100 and 510-amino-acid fragments of the α1 chain of human type I collagen, containing 11 and 55 potential hydroxylation sites, respectively. Contrary to the suggested model, the chain length of the peptide substrate was found to have an identical effect on both the enzyme tetramer and the dimer, as their K_m values for the 510-amino-acid fragment were 0.12% of those for $(Pro\text{-}Pro\text{-}Gly)_5$, and 5-5.6% of those for $(Pro\text{-}Pro\text{-}Gly)_{10}$ (table I in III). All the K_m values of the PHY-1/human PDI dimer were found to be 2-3 times those of the PHY-1/PHY-2/$(PDI\text{-}2)_2$ tetramer (table I in III).

5.3.2 Effect of inactivation of one of the two catalytic sites in a C-P4H tetramer

The effect of inactivation of one of the two catalytic sites in the *C. elegans* PHY-1/PHY-2/$(PDI\text{-}2)_2$ tetramer on its enzyme activity was studied by generating mutant recombinant baculoviruses coding for PHY-1($Asp^{407}\rightarrow Asn$) and PHY-2($Asp^{405}\rightarrow Asn$) in which iron-binding had been abolished, and PHY-1($Lys^{486}\rightarrow Ala$) and PHY-2($Lys^{484}\rightarrow Ala$), containing mutations in the residue binding the C5 carboxyl group of 2-oxoglutarate. Analysis of extracts from insect cells infected with the individual mutant viruses showed that, as expected, the mutant PHY-1 and PHY-2 polypeptides formed insoluble aggregates and their efficient extraction required the use of 1% SDS. Expression levels of the mutant PHY-1 and PHY-2 polypeptides were found to be similar to those of the wild-type polypeptides (figure 3 in III).

Mutant *C. elegans* C-P4H mixed tetramers in which one of the PHY subunits had been subject to mutation and the other subunit was of the wild type (figure 4 in III) were expressed in insect cells and analysed by non-denaturing PAGE. All the mutant PHY polypeptides were able to assemble into mixed tetramers with the wild-type polypeptides, a small amount of wild-type or mutant PHY-1/PDI-2 dimer also being detected (figure 5 in III).

The P4H activity of the mutant enzyme tetramers was analyzed using $(Pro\text{-}Pro\text{-}Gly)_{10}$ as a substrate (table II in III). Inactivation of the iron-binding aspartate in either the PHY-1 or the PHY-2 polypeptide led to a distinct reduction in the enzyme activity, the values being 6 and 8% of that of the wild-type tetramer, respectively. Mutation of the 2-oxoglutarate binding lysine in either of the PHY polypeptides also caused a marked decrease in the activity, which was 26-27% of that of the wild-type tetramer. The activities of all the mutant tetramers were thus reduced by more than 50%. Inactivation of one of the iron-binding sites had a more severe effect, however, than inactivation of one of the 2-oxoglutarate-binding sites (table II in III). No difference was observed in the activity values obtained with the enzymes having a mutation in either the PHY-1 or PHY-2 polypeptide (table II in III). None of the mutations affected the peptide substrate binding, as the K_m values for $(Pro\text{-}Pro\text{-}Gly)_{10}$ were identical between the wild-type and mutant enzymes (Table II in III). Mutation of the iron-binding or 2-oxoglutarate-binding site in one of the two catalytic subunits did not affect binding of these cosubstrates to the remaining wild-type subunit, as the K_m values of the mutant enzymes for these cosubstrates were identical to those of the wild-type enzyme (table III in III).

The P4H activity of the mutant enzyme tetramers was also analyzed with saturating concentrations of recombinant 100 and 510-amino-acid non-hydroxylated α1(I) collagen fragments as substrates. The results obtained with these fragments were very similar to those obtained with $(Pro\text{-}Pro\text{-}Gly)_{10}$, in that mutation of the iron-binding aspartate or the 2-oxoglutarate-binding lysine in either of the PHY polypeptides reducing the enzyme activity to 6-10% and 18-34% of that of the wild-type enzyme, respectively (table IV in III).

6 Discussion

6.1 Human, rat and mouse type III C-P4Hs – new members of the prolyl 4-hydroxylase family

C-P4H was long believed to be of one type only, until an isoform of the vertebrate α subunit, α(II), was cloned and characterized from mouse and human sources (Helaakoski *et al.* 1995, Annunen *et al.* 1997). The human genome is now essentially fully sequenced and a vast number of expressed sequence tags are available in the GenBank (Collins *et al*. 2003). Using EST data, we have now identified, cloned and characterized a third human, rat and mouse C-P4H α subunit isoform, α(III), and demonstrated that, like the α(I) and α(II) subunits, it combines with PDI to form an active $[\alpha(III)]_2\beta_2$ tetramer which hydroxylates procollagen chains and collagen-like peptides. The existence of several isoenzymes seems to be common to many collagen processing enzymes, as three isoforms of lysyl hydroxylase, the enzyme that catalyzes hydroxylation of some Y position lysines in -X-Lys-Gly- collagen repeats, five isoforms of lysyl oxidase, which initiates crosslink formation between collagen molecules in a collagen fibre, and three isoenzymes of the N and C proteinases that cleave the propeptides from procollagen molecules are currently known (see Myllyharju & Kivirikko 2003).

The initial attempts to express recombinant human type III C-P4H were carried out in insect cells, because the two previously known human C-P4Hs and many other P4Hs have been successfully produced in these cells (Vuori *et al.*1992b, Veijola *et al.* 1994, Helaakoski *et al.* 1995, Annunen *et al.* 1997, 1999, Eriksson *et al.* 1999, Hieta & Myllyharju 2002, Riihimaa *et al.* 2002, Hirsilä *et al.* 2003, Winter *et al.* 2003). However, no assembly of the recombinant human α(III) and PDI polypeptides into an active type III C-P4H was obtained. N-terminal sequencing showed that the α(III) signal peptide was not cleaved, and the baculoviral signal peptide GP67, which was also tested, was incorrectly cleaved. The lack of assembly in insect cells thus most probably resulted from inefficient or non-existent transport of the α(III) polypeptide into the ER. An active type III C-P4H tetramer could be expressed in mammalian cells, however, by simultaneously coexpressing recombinant α(III) and human PDI polypeptides. Immunoprecipitation

studies showed that the α(III) and PDI polypeptides are present in the same molecule, and gel filtration studies showed that the elution position of P4H activity from the cells coexpressing these two polypeptides was identical to that of the purified recombinant human type I C-P4H tetramer. Furthermore, non-denaturing PAGE analysis of the gel filtration fractions with the highest P4H activity yielded a band with similar mobility to that of the human type I C-P4H that could be stained with antibodies to the α(III) and PDI polypeptides. These data strongly suggest that the type III C-P4H is an $[\alpha(III)_2]\beta_2$ tetramer and thus has a similar subunit composition to the type I and II C-P4Hs.

The nematode *C. elegans* was found to express unique C-P4H forms, including a mixed tetramer consisting of two distinct types of α subunit and two identical PDI subunits and two forms of αβ dimer (see below), whereas the C-P4H of the filarial nematode *B. malayi* is a homotetramer consisting of four identical α subunits (Winter *et al*. 2003). A C-P4H characterized from another invertebrate, *D. melanogaster*, is an $\alpha_2\beta_2$ tetramer with two identical α subunits (Annunen *et al*. 1999). The P4Hs of plants, algae and viruses and the HIF-P4Hs are evidently monomers (Kaska *et al*. 1987, 1988, Wojtaszek *et al*. 1999, Eriksson *et al*. 1999, Bruick & McKnight 2001, Epstein *et al*. 2001, Hieta & Myllyharju 2002, Ivan *et al*. 2002). The various molecular compositions of the known members of the P4H family are shown in Figure 2. Insect cell coexpression experiments have demonstrated that the human α(I) and α(II) subunits do not assemble into a mixed P4H tetramer with the PDI polypeptide (Annunen *et al*. 1997). Our data do not exclude the possibility that the α(III) polypeptide could form a mixed tetramer with another known or yet to be identified α subunit, but the α(III) subunit does not resemble the *C. elegans* PHY-1 or PHY-2 subunits, as it can form a tetramer with two identical α subunits and does not form an active αβ dimer.

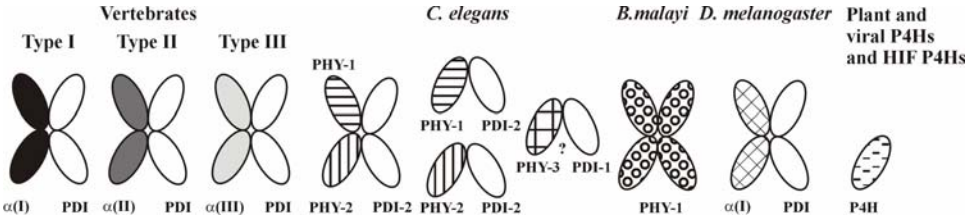


Fig. 2. Forms of prolyl 4-hydroxylase characterized in different species.

The processed vertebrate C-P4H α subunits are very similar in size, the 525-amino-acid human α(III) subunit being slightly longer than the 517 and 514-amino-acid human α(I) and α(II) subunits. The overall amino acid sequence identity between the α(I) and α(II), 65%, is markedly higher than that of 35-37% between α(III) and α(I) or α(II). The amino acid sequences of the individual isoforms are highly conserved between the vertebrate species, the identity between the human and mouse α(I) subunits being 94% (Helaakoski *et al*. 1995), and that between the corresponding α(II) subunits 93% (Annunen *et al*. 1997), while the human α(III) sequence is 91% identical to the rat sequence and 94% identical to that in the mouse. The identity between the same α subunit isoform in different species is thus notably higher than that between the different isoforms in the same species.

The highest identity between the various α subunits is found within the catalytically important C-terminal region, all catalytically critical residues being conserved. The K_m values of the recombinant human type III C-P4H for the cosubstrates Fe^{2+}, 2-oxoglutarate and ascorbate were found to be essentially identical to those of the type I and type II C-P4Hs. The K_i value of the type III C-P4H for the 2-oxoglutarate analogue pyridine 2,4-dicarboxylate was also very similar to those of the other two human C-P4H isoenzymes, while that for pyridine 2,5-dicarboxylate was somewhat higher, this difference possibly being due to the effects of some of the non-conserved residues. These data and the conservation of the catalytically critical residues suggest that the catalytic sites of all three vertebrate C-P4H isoenzymes are highly similar, but the three-dimensional structure of the C-P4H has not yet been resolved. The C-P4Hs belong to the family of 2-oxoglutarate dioxygenases, and the crystal structures of six other members of this family have been resolved, namely deacetoxycephalosporin C synthase, anthocyanidin synthase, proline 3-hydroxylase, clavaminate synthase, taurine/α-ketoglutarate dioxygenase (see Myllyharju 2003) and HIF asparaginyl hydroxylase (Elkins *et al*. 2003, Lee *et al*. 2003). The amino acid sequence identity between these six enzymes, and also a related enzyme isopenicillin *N* synthase which does not utilize 2-oxoglutarate, is very low, but their catalytic domains all adopt a jelly-roll fold formed by eight β strands (Elkins *et al*. 2003, Lee *et al*. 2003, Myllyharju 2003). The Fe^{2+} is co-ordinated by two conserved histidines and one aspartate, and in the 2-oxoglutarate-dependent enzymes the C5 carboxyl group of this cosubstrate forms a salt bridge with a conserved positively charged residue, an arginine, in position +10 with respect to the second iron-binding histidine (see Myllyharju 2003). The HIF asparaginyl hydroxylase is the only exception, in that the 2-oxoglutarate-binding positively charged residue, a lysine, is in position +13 with respect to the iron-binding aspartate (Elkins *et al*. 2003, Lee *et al*. 2003). It is therefore likely that the catalytic domain of the C-P4Hs also contains a jelly-roll core.

The K_m value for the synthetic peptide substrates $(Pro\text{-}Pro\text{-}Gly)_{10}$ and the K_i value for the competitive inhibitor poly(L-proline) of the type III C-P4H were found to lie between those of the type I and II enzymes, being closer to the type I values. The type I enzyme is effectively inhibited by poly(L-proline), whereas the K_i value of the type II enzyme for this peptide is much higher. On the other hand, the type I C-P4H binds $(Pro\text{-}Pro\text{-}Gly)_{10}$ 6 times more effectively than the type II enzyme. It has been shown that the binding properties of various peptides to the type I and type II C-P4Hs are determined by their affinities for a peptide-substrate-binding domain of about 100 amino acids present in the catalytic α subunit (Hieta *et al*. 2003). The differences in peptide-substrate binding properties between the three C-P4H isoenzymes are most probably due to differences in certain amino acid positions in these domains. It has been shown that the presence of a glutamate and glutamine in the α(II) subunit in positions corresponding to the residues Ile182 and Tyr233 in the α(I) subunit explains most of the lack of poly(L-proline) binding in the former (Myllyharju & Kivirikko 1999). Conversion of the residues Ile182 and Tyr233 to glutamate and glutamine, respectively, markedly reduced the binding of poly(L-proline) to the type I enzyme, while conversion of the glutamate and glutamine residues in α(II) to isoleucine and tyrosine, respectively, brought the peptide-substrate and poly(L-proline) binding properties of the type II C-P4H close to those of the type I enzyme (Myllyharju & Kivirikko 1999). The α(III) subunit contains Trp199 and Arg253 in the corresponding positions. The tryptophan residue may be more favourable for

peptide-binding than the negatively charged glutamate in the α(II) subunit, and this may at least partly explain the intermediate peptide binding properties of the type III C-P4H. The α(I) residues Ile182 and Tyr233 and the corresponding glutamine and glutamate residues in the α(II) subunit do not explain all of the peptide binding differences between the two types of C-P4H, however (Myllyharju & Kivirikko 1999). Furthermore, NMR studies have shown that binding of the peptide $(Pro\text{-}Pro\text{-}Gly)_2$ to the type I peptide-substrate-binding domain causes major chemical shifts in the backbone amide resonances of many residues, the residues showing the largest shifts being mainly hydrophobic (Hieta *et al*. 2003). It is possible, therefore, that several amino acid residues may be involved in the determination of the binding properties of the type I, II and III peptide-substrate-binding domains.

The human, rat and mouse α(III) subunits, like the human and mouse α(I) and α(II) subunits, contain two potential attachment sites for asparagine-linked oligosaccharides. The positions of these sites are not conserved in any of the three types of α subunit, however. It has been shown that conversion of the α(I) subunit asparagines at the two glycosylation sites to glutamine has no effect on the assembly of the type I C-P4H tetramer or its P4H activity (Lamberg *et al*. 1995). We showed here that removal of the N-linked oligosaccharides from the α(III) polypeptide likewise had no influence on the P4H activity of the type III enzyme. It is thus obvious that glycosylation of the α subunit isoforms plays no role in their C-P4H function.

The human α(III) polypeptide contains six cysteine residues, while the human α(I) and α(II) isoforms contain 5 conserved cysteines in positions identical to those of cysteines 1 and 3-6 in α(III). The α(II) subunit contains an additional cysteine between the α(III) cysteines 5 and 6. It has been demonstrated by site-directed mutagenesis of the human α(I) subunit that the conserved cysteines corresponding to the α(III) cysteines 3 and 4, and 5 and 6, form intrachain disulphide bonds that are essential for the α(I) subunit to maintain the native structure needed for tetramer assembly, while mutation of the residue corresponding to the α(III) cysteine 1 has no effect on tetramer assembly or enzyme activity (John & Bulleid 1994, Lamberg *et al*. 1995). The α(III) cysteines 1 (Cys165) and 2 (Cys226) are located in a region corresponding to the peptide-substrate-binding domain of the α(I) subunit. NMR studies have shown that the type I domain consists of five α helices that can also be accurately predicted based on its amino acid sequence (Hieta *et al*. 2003). These five α helices can be predicted to be located in identical positions in the α(II) and α(III) polypeptides as well (figure 1 in I). The α(III) cysteines 165 and 226 are located in helices 1 and 4, respectively, and the recently resolved crystal structure of the peptide-substrate-binding domain of the α(I) subunit (Pekkala *et al*. 2003, M. Pekkala, R. Hieta, J. Myllyharju, K. I. Kivirikko & Rik Wierenga, unpublished observations) shows that these two helices are located some distance away from each other, so that it is highly likely that these two α(III) cysteines do not participate in any intrachain disulphide bond formation. Involvement of the α(III) cysteine 226 in interchain disulphide bond formation was excluded here by analysing the soluble fraction of cells coexpressing recombinant human α(III) and PDI polypeptides in SDS-PAGE under reducing and non-reducing conditions. The identical mobility of the α(III) polypeptide under both conditions rules out the existence of any interchain disulphide bonds. This further supports the previous observation that no interchain disulphide bonds exist in a C-P4H tetramer (Nietfeld *et al*. 1981).

The exon-intron organization of the human α(III) subunit gene was found to be very similar to those of the α(I) (Helaakoski *et al*. 1994) and α(II) genes (Nokelainen *et al*. 2001), although the α(III) gene does not contain an exon corresponding to the first non-coding exon of the α(I) and α(II) genes, and only one α(III) exon, number 1, corresponds to exons 2 and 3 in the α(I) and α(II) genes (figure 3A in I). The subsequent exon-intron boundaries are mostly conserved, except that the α(III) exon 3, corresponding to exon 5 in α(I) and α(II), is 86 bp longer, and the region corresponding to exon 6 in α(I) and α(II) is split into two exons in the α(III) gene, while the α(III) sequence corresponding to exons 11 and 12 in α(I) and exons 10 and 11 in α(II) is fused into one exon (figure 3A in I). Minor differences were also found in the lengths of some of the exons. It is probable that the three α subunit genes originate from gene duplications, the α(III) gene resulting from the first duplication, as its coding sequence is least conserved between the three characterized vertebrate isoforms. The α(I) and α(II) genes most probably originate from a more recent duplication, as their coding sequences are highly homologous.

The α(I) and α(II) genes both contain two alternatively spliced exons, but not in conserved positions, being numbers 9 and 10 in the α(I) gene and numbers 12a and 12b in the α(II) gene (Helaakoski *et al*. 1994, Nokelainen *et al*. 2001). Unlike these genes, the α(III) gene does not contain any alternatively spliced exons. The functional significance of the alternatively spliced exons in the α(I) and α(II) genes is still unknown, as α(I) and α(II) subunits encoded by both splicing variants assemble into fully active C-P4H tetramers (Vuori *et al*. 1992b, Nokelainen *et al*. 2001) and as no essential differences have been found in the expression patterns of the alternatively spliced mRNAs between tissues or cell types (Helaakoski *et al*. 1989, 1994, Nokelainen *et al*. 2001).

The human α(III) mRNA was found to be expressed in many tissues, but interestingly, at much lower levels than the α(I) and α(II) mRNAs. The highest expression levels were found in the placenta, adult liver and foetal skin, kidney, liver and lung. It is possible, however, that the α(III) mRNA is expressed at higher levels in certain cells or tissues that were not included in this study. It has been shown by immunohistochemistry and P4H activity measurements that type I C-P4H is the major form in most cell types and tissues, but type II is the main form in cultured chondrocytes and in cartilage, capillary endothelial cells and the developing glomeruli of the foetal kidney (Annunen *et al.* 1998, Nissi *et al.* 2001). The α(III) subunit has also been cloned and characterized recently by another independent group (Van Den Diepstraten *et al*. 2003), who have shown by *in situ* hybridization that the α(III) mRNA is expressed in atherosclerotic carotid artery samples. The strongest expression was detected in smooth muscle cell-like cells within the fibrous caps of lipid-rich lesions, whereas little if any expression was seen in the media of non-atherosclerotic artery samples. Expression of the α(I) and α(II) mRNAs was not studied in the atherosclerotic artery samples, but all three subunit mRNAs were shown to be expressed in human internal thoracic B5 smooth muscle cells, although only the α(II) mRNA showed an increased expression level during maturation (Van Den Diepstraten *et al*. 2003). It has also been shown previously in an immunofluorescence study that the α(I) subunit is expressed in the smooth muscle cells of the large arteries (Nissi *et al*. 2001). Knock-out mice for each of the three vertebrate C-P4H α subunits identified so far will be very valuable for the identification of possible specific roles and expression patterns of the C-P4H isoenzymes.

6.2 A unique PHY-1/PHY-2/$(PDI-2)_2$ mixed tetramer is the main C-P4H form in wild-type *C. elegans*

We demonstrated here that the *C. elegans* polypeptides PHY-1, PHY-2 and PDI-2 assemble into three novel combinations of subunit: a PHY-1/PHY-2/$(PDI-2)_2$ mixed tetramer and PHY-1/PDI-2 and PHY-2/PDI-2 dimers, of which only the dimer PHY-1/PDI-2 has been characterized previously (Veijola *et al*. 1994, 1996). All three active C-P4H forms are unique, because the vertebrate α(I) and α(II) subunits do not form a mixed tetramer with PDI (Annunen *et al*. 1997) and no αβ dimers have so far been reported in any other species. We showed here that formation of the *C. elegans* mixed tetramer is highly species-specific, as none of the subunits of the PHY-1/PHY-2/$(PDI-2)_2$ tetramer could be replaced by the human or *Drosophila* C-P4H α or β subunits. Thus various molecular compositions are utilized in the P4H family members characterized so far (Figure 2), which include the monomeric plant, algal and viral P4Hs and HIF-P4Hs, the C-P4Hs in which PDI serves as a β subunit, i.e. the dimers and the mixed tetramer in *C. elegans,* and the vertebrate and *Drosophila* C-P4H tetramers, each containing a single type of the catalytic subunit, and even a homotetramer C-P4H in the nematode *B. malayi* (Helaakoski *et al*. 1989, 1995, Veijola *et al*. 1994, Annunen *et al*. 1997, 1999, Eriksson *et al*. 1999, Bruick & McKnight 2001, Epstein *et al*. 2001, Hieta & Myllyharju 2002, Ivan *et al*. 2002, Winter *et al*. 2003).

The determinants of the molecular composition of a given P4H are currently unknown, but the *C. elegans* mixed tetramer did offer a novel possibility to identify regions in the two PHY polypeptides that are critical for its assembly. We generated four recombinant hybrid PHY polypeptides and studied their assembly into a mixed tetramer with a wild-type PHY-1 or PHY-2 or with PDI-2 (figure 7 in II). The most obvious difference between the two PHY polypeptides is the C-terminal extension of 18 amino acids present in PHY-1, but this is not involved in tetramer formation, as mixed tetramers were successfully assembled from PHY polypeptides that had either a PHY-1 or a PHY-2 C-terminal region. Instead, the critical amino acids were shown to be located in different regions in the two polypeptides, between the PHY-1 residues Gln121 and Ala271 and between the PHY-2 residues Asp1 and Leu122. The amino acid sequence identity between the PHY-2 residues 1-122 and the corresponding region of PHY-1 is 45% and that between residues 121-271 of PHY-1 and the corresponding region of PHY-2 is 47%, whereas higher amino acid sequence identity, 60%, is found between the 271 C-terminal residues of PHY-1 and PHY-2.

The identification of the novel C-P4H forms in *C. elegans* offers for the first time an explanation for many previous genetic observations made with respect to this organism. It has been shown that inactivation of the *phy-2* gene causes no visible phenotype, while deletion of the *phy-1* gene results in a short and fat, dumpy phenotype (Friedman *et al*. 2000, Hill *et al*. 2000, Winter & Page 2000) and combined disruption of the *phy-1* and *phy-2* genes leads to embryonic lethality (Friedman *et al*. 2000, Winter & Page 2000). Our analyses of the nematode lysates showed that the PHY-1/PHY-2/$(PDI-2)_2$ mixed tetramer is the main C-P4H form in wild-type *C. elegans in vivo*, the PHY-1/PDI-2 dimer also being present, but at a much lower level. This finding agrees with our insect cell data showing that assembly of the recombinant PHY-1/PDI-2 dimer is quite inefficient and the

level of P4H activity generated in cells coexpressing these two polypeptides is only about 7% of that generated in cells expressing the mixed tetramer. The mixed enzyme tetramer was not formed in the phenotypically wild-type *phy-2* mutant strain, because the PHY-2 polypeptide was lacking, but assembly of the PHY-1/PDI-2 dimer was greatly enhanced and about 54-57% of the P4H activity found in the wild-type worms was still retained. The 4-hydroxyproline to proline ratio in *phy-2* null nematode cuticles has been shown to be reduced to 0.829 from the figure of 1.297 observed in wild-type cuticles (Friedman *et al*. 2000). Although not leading to complete hydroxylation of the cuticle collagens, increased assembly of the PHY-1/PDI-2 dimer can obviously compensate for the lack of the mixed tetramer, in that the *phy-2* mutant nematodes are phenotypically of the wild type.

The dumpy *phy-1* null nematodes showed abolition of the assembly of both the mixed tetramer and the PHY-1/PDI-2 dimer, whereas a novel PHY-2/PDI-2 dimer was detected. Although the assembly of this PHY-2/PDI-2 dimer is very inefficient, its P4H activity level being only 0.7–2.5% of that in wild-type extracts, it may partially compensate for the lack of the other two forms. The *phy-1* mutants survive, but the marked reduction in the 4-hydroxyproline content of their cuticle collagens (Friedman *et al*., Winter & Page 2000) leads to the dumpy phenotype. Inactivation of both the *phy-1* and *phy-2* genes simultaneously (Friedman *et al*. 2000, Winter & Page 2000) or inactivation of the *pdi-2* gene (Winter & Page 2000) leads to embryonic lethality, as the assembly of all three C-P4H forms is abolished.

We also showed that the PHY-1, PHY-2 and PDI-2 polypeptides are coexpressed in the hypodermal cells synthesizing cuticle collagen. It has previously been shown that the maximal synthesis of these polypeptides coincides with the maximal synthesis of collagens during moulting (Winter & Page 2000). A third *C. elegans phy* gene, *phy-3*, has recently been cloned and characterized, and the PHY-3 polypeptide has been shown to be involved in the hydroxylation of egg shell collagens (Riihimaa *et al*. 2002). We showed here that assembly of the PHY-1/PHY-2/$(PDI\text{-}2)_2$ tetramer and the PHY-1/PDI-2 dimer in the *phy-3* null nematodes is similar to that in the wild-type nematodes. RNAi studies of the *phy-3* transcript also demonstrated that PHY-3 has no role in the assembly of the C-P4H forms that are involved in the synthesis of cuticle collagens. The data obtained here show that PHY-1 and PHY-2 represent the only catalytic C-P4H subunits needed for hydroxylation of cuticular collagens and that the presence of these subunits is critical for normal growth and body shape in the nematode *C. elegans*.

The recombinant *C. elegans* mixed C-P4H tetramer had very similar catalytic properties to the human C-P4Hs and the previously studied *C. elegans* PHY-1/human PDI dimer, but it was not inhibited by poly(L-proline) at low concentrations and thus resembled more the human type II C-P4H. It was also inhibited efficiently by 2-oxoglutarate analogues that are known to be efficient inhibitors of the vertebrate C-P4Hs. We studied the *in vivo* effects of two such compounds on live nematodes. A limited range of effects was detected in the wild-type and *phy-2* mutant nematodes, whereas severe temperature-dependent effects were noted in the *phy-1* mutant strain. Such inhibitors could thus be regarded as potential targets for the development of drugs to control parasitic nematode species.

6.3 Roles of the two catalytic sites in a C-P4H tetramer

The possibility of co-operation between the two catalytic α subunits of C-P4H tetramers in the hydroxylation of proline residues has been studied previously using the vertebrate C-P4H with two identical α subunits as a model (de Waal & de Jong 1988, de Jong *et al*. 1991). Based on the obtained results a theory of a processive mechanism of binding has been suggested according to which the two peptide-binding sites on one enzyme molecule work co-operatively, so that the K_m values decrease when hydroxylating long peptide substrates. This theory maintains that the effect of peptide substrate length would not be observed with an enzyme having only one peptide-binding site and one catalytic site. The unique C. *elegans* C-P4H forms, i.e. the mixed tetramer and the dimer, offer convenient tools for studying the hydroxylation mechanism and the roles of the individual catalytic subunits further.

We determined here K_m values for four peptide substrates of various lengths using the recombinant *C. elegans* mixed tetramer and the hybrid *C. elegans* PHY-1 /human PDI dimer. The latter was chosen because its assembly in insect cells is much more efficient than that of the PHY-1/*C.elegans* PDI-2 dimer. Interestingly, the marked decrease in the K_m values with increasing peptide length was seen both with the enzyme tetramer, having two peptide-binding and catalytic sites, and with the dimer, having only one peptide-binding and catalytic site. The data on the catalytic properties of the monomeric P4Hs from the plant *Arabidopsis thaliana* and the *Paramecium bursaria Chlorella* virus-1 (Eriksson *et al*. 1999, Hieta & Myllyharju 2002) are in agreement with our results, as a similar decrease in the K_m values for long peptide substrates was observed with them. In contrast, the partially characterized algal P4H monomer from *Chamydomonas reinhardtii* has been reported to hydroxylate long poly-(L-proline) peptides no more efficiently than short ones (Kaska *et al*. 1987).

It has recently been shown that the single recombinant peptide-substrate-binding domains of the human type I and II C-P4Hs bind longer peptide substrates more efficiently than shorter ones (Hieta *et al*. 2003), the K_d values of these domains for various peptides being very similar to the K_m and K_i values of the enzyme tetramers (Hieta *et al*. 2003). Our data further support the idea that this peptide length dependence most probably results from more effective binding of longer peptides to the peptide-substrate-binding domain rather than a processive action involving two binding sites. The *A. thaliana* and viral P4Hs do not have amino acid sequences similar to that of the peptide-substrate-binding domain located in the N-terminal half of the catalytic subunit of vertebrate and *C. elegans* C-P4Hs (Eriksson *et al*. 1999, Hieta & Myllyharju 2002). Furthermore, these enzymes hydroxylate many proline-rich peptides, and as poly(L-proline), which are not used as substrates for the vertebrate and *C. elegans* C-P4Hs (Eriksson *et al*. 1999, Hieta & Myllyharju 2002). It is thus probable that binding of the peptide substrates occurs in a different manner in the monomeric P4Hs.

The individual roles of the two catalytic sites in a C-P4H tetramer were studied by mutations of a critical iron-binding or 2-oxoglutarate-binding amino acid in one of the catalytic subunits in the *C. elegans* PHY-1/PHY-2/$(PDI\text{-}2)_2$ mixed tetramer. Surprisingly, inactivation of one catalytic subunit reduced the P4H activity by more than 50%, which suggests that the wild-type subunit cannot function fully independently. The results did

show, however, that the catalytic sites in both subunits are functional and have virtually equal roles, as essentially identical activity values were obtained with enzymes having a mutation in either the PHY-1 or the PHY-2 polypeptide. The fact that the K_m values of the mutant enzymes for Fe^{2+} and 2-oxoglutarate were identical to those of the wild-type enzyme suggests that the more than 50% reduction in the activity of the mutant tetramers is not due to impaired cosubstrate binding to the wild-type catalytic subunit.

It has been demonstrated previously that the K_d value of the recombinant human type I peptide-substrate-binding domain for a synthetic peptide substrate that contains 4-hydroxyproline is more than one order of magnitude higher than that for a non-hydroxylated peptide (Hieta *et al*. 2003). It is possible that when one catalytic site is inactivated, the non-hydroxylated peptide bound to the peptide-substrate-binding domain is not released or is released more slowly, and thus interferes with the interaction between the functional α subunit and the peptide substrate.

One unexpected finding was that mutation of the Fe^{2+}-binding residue in either of the PHY polypeptides led to a 4 times more severe decrease in P4H activity than mutation of the 2-oxoglutarate-binding residue. According to previous data, mutation of either the iron-binding aspartate or the 2-oxoglutarate-binding positively charged residue at both catalytic sites of the human type I C-P4H tetramer or of lysyl hydroxylase, a homodimer, or at the single catalytic site of the *A. thaliana* P4H completely inactivates these enzymes (Pirskanen *et al*. 1996, Myllyharju & Kivirikko 1997, Passoja *et al*.1998, Hieta & Myllyharju 2002).

References

Annunen P, Autio-Harmainen H & Kivirikko KI (1998) The novel type II prolyl 4-hydroxylase is the main enzyme form in chondrocytes and capillary endothelial cells, whereas type I enzyme predominates in most cells. J Biol Chem 273: 5989-5992.

Annunen P, Helaakoski T, Myllyharju J, Veijola J, Pihlajaniemi T & Kivirikko KI (1997) Cloning of the human prolyl 4-hydroxylase α subunit isoform α(II) and characterization of the type II enzyme tetramer. The α(I) and α(II) subunits do not form a mixed α(I)α(II)β_2 tetramer. J Biol Chem 272: 17342-17348.

Annunen P, Koivunen P & Kivirikko KI (1999) Cloning of the α subunit of prolyl 4-hydroxylase from *Drosophila* and expression and characterization of the corresponding enzyme tetramer with some unique properties. J Biol Chem 274: 6790-6796.

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

Bella J, Brodsky B & Berman HM (1995) Hydration structure of a collagen peptide. Structure 3: 893-906.

Berg RA & Prockop DJ (1973) The thermal transition of a non-hydroxylated form of collagen. Evidence for a role for hydroxyproline in stabilizing the triple-helix of collagen. Biochem Biophys Res Commun 52: 115-120.

Bolwell GP, Robbins MP & Dixon RA (1985) Elicitor-induced prolyl hydroxylase from French bean (*Phaseolus vulgaris*). Localization, purification and properties. Biochem J 229:693-699.

Bruick RK & McKnight SL (2001) A conserved family of prolyl-4-hydroxylases that modify HIF. Science 294: 1337-1340.

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.

Collins FS, Green ED, Guttmacher AE & Guyer MS; US National Human Genome Research Institute (2003) A vision for the future of genomics research. Nature 422: 835-47.

Crossen R & Gruenwald S (1998) Baculovirus Expression Vector System Instruction Manual, PharMingen, San Diego, CA.

Culetto E & Sattelle DB (2000) A role for *Caenorhabditis elegans* in understanding the function and interactions of human disease genes. *Hum Mol Genet* 9: 869-877.

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

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

Dijkstra K, Karvonen P, Pirneskoski A, Koivunen P, Kivirikko KI, Darby NJ, van Straaten M, Scheek RM & Kemmink J (1999) Assignment of ^{1}H, ^{13}C and ^{15}N resonances of the *a'* domain of protein disulfide isomerase. J Biomol NMR 14: 195-196.

Edman JC, Ellis L, Blacher RW, Roth RA & Rutter WJ (1985) Sequence of protein disulphide isomerase and implications of its relationship to thioredoxin. Nature 317: 267-270.

Elkins JM, Hewitson KS, McNeill LA, Seibel JF, Schlemminger I, Pugh CW, Ratcliffe PJ & Schofield CJ (2003) Structure of factor-inhibiting hypoxia-inducible factor (HIF) reveals mechanism of oxidative modification of HIF-1α. J Biol Chem 278: 1802-1806.

Epstein AC, Gleadle JM, McNeill LA, Hewitson KS, O'Rourke J, Mole DR, Mukherji M, Metzen E, Wilson MI, Dhanda A, Tian Y-M, Masson N, Hamilton DL, Jaakkola P, Barstead R, Hodgkin J, Maxwell PH, Pugh CW, Schofield CJ & Ratcliffe PJ (2001) *C. elegans* EGL-9 and mammalian homologs define a family of dioxygenases that regulate HIF by prolyl hydroxylation. Cell 107: 43-54.

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

Fedele AO, Whitelaw ML & Peet DJ (2002) Regulation of gene expression by the hypoxia-inducible factors. Mol Intervent 2: 229-243.

Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE & Mello CC (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. Nature 391:806-811.

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

Guo XD, Johnson JJ & Kramer JM (1991) Embryonic lethality caused by mutations in basement membrane collagen of *C. elegans*. Nature 349: 707-709.

Gupta MC, Graham PL & Kramer JM (1997) Characterization of α1(IV) collagen mutations in *Caenorhabditis elegans* and the effects of α1 and α2(IV) mutations on type IV collagen distribution. J Cell Biol 137: 1185-1196.

Hanauske-Abel HM & Günzler V (1982) A stereochemical concept for the catalytic mechanism of prolylhydroxylase: applicability to classification and design of inhibitors. J Theor Biol 94: 421-455.

Hariharan IK & Haber DA (2003) Yeast, flies, worms, and fish in the study of human disease. New Engl J Med 348: 2457-2463.

Hawkins HC & Freedman RB (1991) The reactivities and ionization properties of the active-site dithiol groups of mammalian protein disulphide-isomerase. Biochem J 275: 335-339.

Helaakoski T, Annunen P, Vuori K, MacNeil IA, Pihlajaniemi T & Kivirikko KI (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, Veijola J, Vuori K, Rehn M, Chow LT, Taillon-Miller P, Kivirikko KI & Pihlajaniemi T (1994) Structure and expression of the human gene for the α subunit of prolyl 4-hydroxylase. The two alternatively spliced types of mRNA correspond to two homologous exons the sequences of which are expressed in a variety of tissues. J Biol Chem 269: 27847-27854.

Helaakoski T, Vuori K, Myllylä R, Kivirikko KI & 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.

Hewitson KS, McNeill LA, Riordan MV, Tian Y-M, Bullock AN, Welford RW, Elkins JM, Oldham NJ, Bhattacharya S, Gleadle JM, Ratcliffe PJ, Pugh CW & Schofield CJ (2002) Hypoxia-inducible factor (HIF) asparagine hydroxylase is identical to factor inhibiting HIF (FIH) and is related to the cupin structural family. J Biol Chem 277: 26351-26355.

Hieta R, Kukkola L, Permi P, Pirilä P, Kivirikko KI, Kilpeläinen 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

Hieta R & Myllyharju J (2002) Cloning and characterization of a low molecular weight prolyl 4-hydroxylase from *Arabidopsis thaliana*. Effective hydroxylation of proline-rich, collagen-like, and hypoxia-inducible transcription factor α-like peptides. J Biol Chem 277: 23965-23971.

Hill KL, Harfe BD, Dobbins CA & L'Hernault SW (2000) *dpy-18* encodes an α-subunit of prolyl-4-hydroxylase in *Caenorhabditis elegans*. Genetics 155: 1139-1148.

Hirsilä M, Koivunen P, Günzler V, Kivirikko KI & Myllyharju J (2003) Characterization of the human prolyl 4-hydroxylases that modify the hypoxia-inducible factor HIF. J Biol Chem 278: 30772-30780.

Holmgren SK, Bretscher LE, Taylor KM & Raines RT (1999) A hyperstable collagen mimic. Chem Biol 6: 63-70.

Holmgren SK, Taylor KM, Bretscher LE & Raines RT (1998) Code for collagen's stability deciphered. Nature 392: 666-667.

Ivan M, Haberberger T, Gervasi DC, Michelson KS, Günzler V, Kondo K, Yang H, Sorokina I, Conaway RC, Conaway JW & Kaelin WG Jr (2002) Biochemical purification and pharmacological inhibition of a mammalian prolyl hydroxylase acting on hypoxia-inducible factor. Proc Natl Acad Sci USA 99: 13459-13464.

Ivan M, Kondo K, Yang H, Kim W, Valiando J, Ohh M, Salic A, Asara JM, Lane WS & Kaelin WG Jr (2001) HIFα targeted for VHL-mediated destruction by proline hydroxylation: implications for O_2 sensing. Science 292: 464-468

Jaakkola P, Mole DR, Tian Y-M, Wilson MI, Gielbert J, Gaskell SJ, von Kriegsheim A, Hebestreit HF, Mukherji M, Schofield CJ, Maxwell PH, Pugh CW & Ratcliffe PJ (2001) Targeting of HIF-α to the von Hippel-Lindau ubiquitylation complex by O_2-regulated prolyl hydroxylation. Science 292: 468-472.

Jenkins CL & Raines RT (2002) Insights on the conformational stability of collagen. Nat Prod Rep 19: 49-59.

Jiang F, Peisach J, Ming LJ, Que L Jr & Chen VJ (1991) Electron spin echo envelope modulation studies of the Cu(II)-substituted derivative of isopenicillin N synthase: a structural and spectroscopic model. Biochemistry 30: 11437-11445.

Jiang H, Guo R & Powell-Coffman JA (2001) The *Caenorhabditis elegans hif-1* gene encodes a bHLH-PAS protein that is required for adaptation to hypoxia. Proc Natl Acad Sci USA 98: 7916-7921.

John DC & Bulleid NJ (1994) Prolyl 4-hydroxylase: defective assembly of α-subunit mutants indicates that assembled α-subunits are intramolecularly disulfide bonded. Biochemistry 33: 14018-14025.

Johnstone IL (1994) The cuticle of the nematode *Caenorhabditis elegans*: a complex collagen structure. Bioessays 16: 171-178.

Johnstone IL (2000) Cuticle collagen genes. Expression in *Caenorhabditis elegans*. Trends Genet 16: 21-27.

Kaska DD, Günzler V, Kivirikko KI & Myllylä R (1987) Characterization of a low-relative-molecular-mass prolyl 4-hydroxylase from the green alga *Chlamydomonas reinhardii*. Biochem J 241: 483-490.

Kaska DD, Myllylä R, Günzler V, Gibor A & Kivirikko KI (1988) Prolyl 4-hydroxylase from *Volvox carteri*. A low-M_r enzyme antigenically related to the α subunit of the vertebrate enzyme. Biochem J 256: 257-263.

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

Kielty CM & Grant ME (2002) The collagen family: structure, assembly, and organization in the extracellular matrix In: Royce PM & Steinmann B (eds) Connective Tissue and Its Heritable Disorders, Wiley-Liss, Inc., New York, 159-221.

Kivirikko KI & Myllyharju J (1998) Prolyl 4-hydroxylases and their protein disulfide isomerase subunit. Matrix Biol 16: 357-368.

Kivirikko KI & Myllylä R (1982) Posttranslational enzymes in the biosynthesis of collagen: intracellular enzymes. Methods Enzymol 82: 245-304.

Kivirikko KI, Myllylä R & Pihlajaniemi T (1992) Hydroxylation of proline and lysine residues in collagens and other animal and plant proteins. In: Harding JJ & Crabbe MJC (eds) Post-Translational Modifications of Proteins. CRC Press, Boca Raton, 1-51.

Kivirikko KI & Pihlajaniemi T (1998) Collagen hydroxylases and the protein disulfide isomerase subunit of prolyl 4-hydroxylases. Adv Enzymol Relat Areas Mol Biol 72: 325-398.

Lamandé SR & Bateman JF (1999) Procollagen folding and assembly: the role of endoplasmic reticulum enzymes and molecular chaperones. Semin Cell Dev Biol 10: 455-464.

LaMantia M & Lennarz WJ (1993) The essential function of yeast protein disulfide isomerase does not reside in its isomerase activity. Cell 74: 899-908.

Lamberg A, Pihlajaniemi T & Kivirikko KI (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.

Lando D, Peet DJ, Gorman JJ, Whelan DA, Whitelaw ML & Bruick RK (2002) FIH-1 is an asparaginyl hydroxylase enzyme that regulates the transcriptional activity of hypoxia-inducible factor. Genes Dev 16:1466-1471.

Lawrence CC, Sobey WJ, Field RA, Baldwin JE & Schofield CJ (1996) Purification and initial characterization of proline 4-hydroxylase from *Streptomyces griseoviridus* P8648: a 2-oxoacid, ferrous-dependent dioxygenase involved in etamycin biosynthesis. Biochem J 313: 185-191.

Lebeche D, Lucero HA & Kaminer B (1994) Calcium binding properties of rabbit liver protein disulfide isomerase. Biochem Biophys Res Commun 202: 556-561.

Lee C, Kim SJ, Jeong DG, Lee SM & Ryu SE (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.

Link CD, Taft A, Kapulkin V, Duke K, Kim S, Fei Q, Wood DE & Sahagan BG (2003) Gene expression analysis in a transgenic *Caenorhabditis elegans* Alzheimer's disease model. Neurobiol Aging 24: 397-413.

Mahon PC, Hirota K & Semenza GL (2001) FIH-1: a novel protein that interacts with HIF-1α and VHL to mediate repression of HIF-1 transcriptional activity. Genes Dev 15: 2675-2686.

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

Maxwell PH, Wiesener MS, Chang GW, Clifford SC, Vaux EC, Cockman ME, Wykoff CC, Pugh CW, Maher ER & Ratcliffe PJ (1999) The tumour suppressor protein VHL targets hypoxia-inducible factors for oxygen-dependent proteolysis. Nature 399: 271-275.

Medveczky MM, Geck P, Vassallo R & Medveczky PG (1993) Expression of the collagen-like putative oncoprotein of *Herpesvirus saimiri* in transformed T cells. Virus Genes 7: 349-365.

Merriweather A, Günzler V, Brenner M & Unnasch TR (2001) Characterization and expression of enzymatically active recombinant filarial prolyl 4-hydroxylase. Mol Biochem Parasitol 116: 185-197.

Ming LJ, Que L Jr, Kriauciunas A, Frolik CA & Chen VJ (1991) NMR studies of the active site of isopenicillin N synthase, a non-heme iron(II) enzyme. Biochemistry 30: 11653-11659.

Montgomery MK, Xu S & Fire A (1998) RNA as a target of double-stranded RNA-mediated genetic interference in *Caenorhabditis elegans*. Proc Natl Acad Sci USA 95: 15502-15507.

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

Myllyharju J & Kivirikko KI (1997) Characterization of the iron- and 2-oxoglutarate-binding sites of human prolyl 4-hydroxylase. EMBO J 16: 1173-1180.

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

Myllyharju J & Kivirikko KI (2001) Collagens and collagen-related diseases. Ann Med 33: 7-21.

Myllyharju J & Kivirikko KI (2003) Collagens and their mutations: from man to *Drosophila and Caenorhabditis elegans*. Trends Genet, in press.

Myllylä R, Günzler V, Kivirikko KI & Kaska DD (1992) Modification of vertebrate and algal prolyl 4-hydroxylases and vertebrate lysyl hydroxylase by diethyl pyrocarbonate. Evidence for histidine residues in the catalytic site of 2-oxoglutarate-coupled dioxygenases. Biochem J 286: 923-927.

Myllylä R, Majamaa K, Günzler V, Hanauske-Abel HM & Kivirikko KI (1984) Ascorbate is consumed stoichiometrically in the uncoupled reactions catalyzed by prolyl 4-hydroxylase and lysyl hydroxylase. J Biol Chem 259: 5403-5405.

Myllylä R, Tuderman L & Kivirikko KI (1977) Mechanism of the prolyl hydroxylase reaction. 2. Kinetic analysis of the reaction sequence. Eur J Biochem 80: 349-357.

Nambu JR, Chen W, Hu S & Crews ST (1996) The *Drosophila melanogaster* similar bHLH-PAS gene encodes a protein related to human hypoxia-inducible factor 1 α and *Drosophila* single-minded. Gene 172: 249-254.

Nietfeld JJ, 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.

Nissi R, Autio-Harmainen H, Marttila P, Sormunen R & Kivirikko KI (2001) Prolyl 4-hydroxylase isoenzymes I and II have different expression patterns in several human tissues. J Histochem Cytochem 49: 1143-1153.

Noiva R (1999) Protein disulfide isomerase: the multifunctional redox chaperone of the endoplasmic reticulum. Semin Cell Dev Biol 10: 481-493.

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

Nokelainen M, Nissi R, Kukkola L, Helaakoski T & Myllyharju J (2001) Characterization of the human and mouse genes for the α subunit of type II prolyl 4-hydroxylase. Identification of a previously unknown alternatively spliced exon and its expression in various tissues. Eur J Biochem 268: 5300-5309.

O'Kane CJ (2003) Modelling human diseases in *Drosophila* and *Caenorhabditis*. Semin Cell Dev Biol 14: 3-10.

Otsu M, Omura F, Yoshimori T & Kikuchi M (1994) Protein disulfide isomerase associates with misfolded human lysozyme *in vivo*. J Biol Chem 269: 6874-6877.

Page AP (2001) The nematode cuticle: synthesis, modification and mutants. In: Kennedy MW & Harnett W (eds) Parasitic Nematodes, CABI Press, United Kingdom, 167-193.

Passoja K, Myllyharju J, Pirskanen A & Kivirikko KI (1998) Identification of arginine-700 as the residue that binds the C5 carboxyl group of 2-oxoglutarate in human lysyl hydroxylase 1. FEBS Lett 434: 145-148.

Pekkala M, Hieta R, Kursula P, Kivirikko KI, Wierenga RK & Myllyharju J (2003) Crystallization of the proline-rich-peptide binding domain of human type I collagen prolyl 4-hydroxylase. Acta Cryst D 59: 940-942.

Pelham HR (1990) The retention signal for soluble proteins of the endoplasmic reticulum. Trends Biochem Sci 15: 483-486.

Pihlajaniemi T, Helaakoski T, Tasanen K, Myllylä R, Huhtala ML, Koivu J & Kivirikko KI (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.

Pirskanen A, Kaimio A-M, Myllylä R & Kivirikko KI (1996) Site-directed mutagenesis of human lysyl hydroxylase expressed in insect cells. Identification of histidine residues and an aspartic acid residue critical for catalytic activity. J Biol Chem 271: 9398-9402.

Ramshaw JAM & Brodsky B (2003) The role of hydroxyproline. Progr Biotechnol 23: 139-141.

Randall CR, Zang Y, True AE, Que L Jr, Charnock JM, Garner CD, Fujishima Y, Schofield CJ & Baldwin JE (1993) X-ray absorption studies of the ferrous active site of isopenicillin N synthase and related model complexes. Biochemistry 32: 6664-6673.

Rasmussen M, Jacobsson M & Björck L (2003) Genome-based identification and analysis of collagen-related structural motifs in bacterial and viral proteins. J Biol Chem 278: 32313-32316.

Riihimaa P, Nissi R, Page AP, Winter AD, Keskiaho K, Kivirikko KI & Myllyharju J (2002) Egg shell collagen formation in *Caenorhabditis elegans* involves a novel prolyl 4-hydroxylase expressed in spermatheca and embryos and possessing many unique properties. J Biol Chem 277: 18238-18243.

Roach PL, Clifton IJ, Fülop V, Harlos K, Barton GJ, Hajdu J, Andersson I, Schofield CJ & Baldwin JE (1995) Crystal structure of isopenicillin N synthase is the first from a new structural family of enzymes. Nature 375: 700-704.

Rosenbloom J & Abrams WR (2002) Elastin and the microfibrillar apparatus. In: Royce PM & Steinmann B (eds) Connective Tissue and Its Heritable Disorders, Wiley-Liss, Inc., New York, 249-269.

Rosenbloom J, Harsch M & Jimenez S (1973) Hydroxyproline content determines the denaturation temperature of chick tendon collagen. Arch Biochem Biophys 158: 478-484.

Semenza GL (2000) HIF-1 and human disease: one highly involved factor. Genes Dev. 14: 1983-1991.

Semenza GL (2001) HIF-1, O_2, and the 3 PHDs: how animal cells signal hypoxia to the nucleus. Cell 107: 43-54.

Showalter AM (1993) Structure and function of plant cell wall proteins. Plant Cell 5:9-23.

Smith MC, Burns N, Sayers JR, Sorrell JA, Casjens SR & Hendrix RW (1998) Bacteriophage collagen. Science 279: 1834.

Tabara H, Grishok A & Mello CC (1998) RNAi in *C. elegans*: soaking in the genome sequence. Science 282: 430-431.

Tanaka M, Shibata H & Uchida T (1980) A new prolyl hydroxylase acting on poly-L-proline, from suspension cultured cells of *Vinca rosea*. Biochim Biophys Acta 616: 188-198.

The *C. elegans* Sequencing Consortium (1998) Genome sequence of the nematode *C. elegans*: a platform for investigating biology. Science 282: 2012-2018.

Timmons L, Court DL & Fire A (2001) Ingestion of bacterially expressed dsRNAs can produce specific and potent genetic interference in *Caenorhabditis elegans*. Gene 263: 103-112.

Tuderman L, Myllylä R & Kivirikko KI (1977) Mechanism of the prolyl hydroxylase reaction. 1. Role of co-substrates. Eur J Biochem 80: 341-348.

Van Den Diepstraten C, Papay K, Bolender Z, Brown A & Pickering JG (2003) Cloning of a novel prolyl 4-hydroxylase subunit expressed in the fibrous cap of human atherosclerotic plaque. Circulation 108: 508-511.

Veijola J, Annunen P, Koivunen P, Page AP, Pihlajaniemi T & Kivirikko KI (1996) Baculovirus expression of two protein disulphide isomerase isoforms from *Caenorhabditis elegans* and characterization of prolyl 4-hydroxylases containing one of these polypeptides as their β subunit. Biochem J 317: 721-729.

Veijola J, Koivunen P, Annunen P, Pihlajaniemi T & Kivirikko KI (1994) Cloning, baculovirus expression, and characterization of the α subunit of prolyl 4-hydroxylase from the nematode *Caenorhabditis elegans*. This α subunit forms an active αβ dimer with the human protein disulfide isomerase/β subunit. J Biol Chem 269: 26746-26753.

Vuori K, Myllylä R, Pihlajaniemi T & Kivirikko KI (1992a) Expression and site-directed mutagenesis of human protein disulfide-isomerase in *Escherichia coli*. This multifunctional polypeptide has two independently acting catalytic sites for the isomerase activity. J Biol Chem 267: 7211-7214.

Vuori K, Pihlajaniemi T, Marttila M & Kivirikko KI (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 KI (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.

Wang GL, Jiang BH, Rue EA & Semenza GL (1995) Hypoxia-inducible factor 1 is a basic-helix-loop-helix-PAS heterodimer regulated by cellular O_2 tension. Proc Natl Acad Sci USA 92: 5510-5514.

Wetterau JR, Combs KA, McLean LR, Spinner SN & Aggerbeck LP (1991) Protein disulfide isomerase appears necessary to maintain the catalytically active structure of the microsomal triglyceride transfer protein. Biochemistry 30: 9728-97235.

Winter AD, Myllyharju J & Page AP (2003) A hypodermally expressed prolyl 4-hydroxylase from the filarial nematode *Brugia malayi* is soluble and active in the absence of protein disulfide isomerase. J Biol Chem 278: 2554-2562.

Winter AD & Page AP (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.

Wojtaszek P, Smith CG & Bolwell GP (1999) Ultrastructural localisation and further biochemical characterisation of prolyl 4-hydroxylase from *Phaseolus vulgaris*: comparative analysis. Int J Biochem Cell Biol 31: 463-477.

Wood WB (1988) Embryology. In: Wood WB (ed) The nematode *Caenorhabditis elegans*. Cold Spring Harbor Laboratory Press, Plainview, NY, 215-241.

World Health Organization (2000) Lymphatic filariasis. WHO Monogr Ser Fact Sheet no 102.