

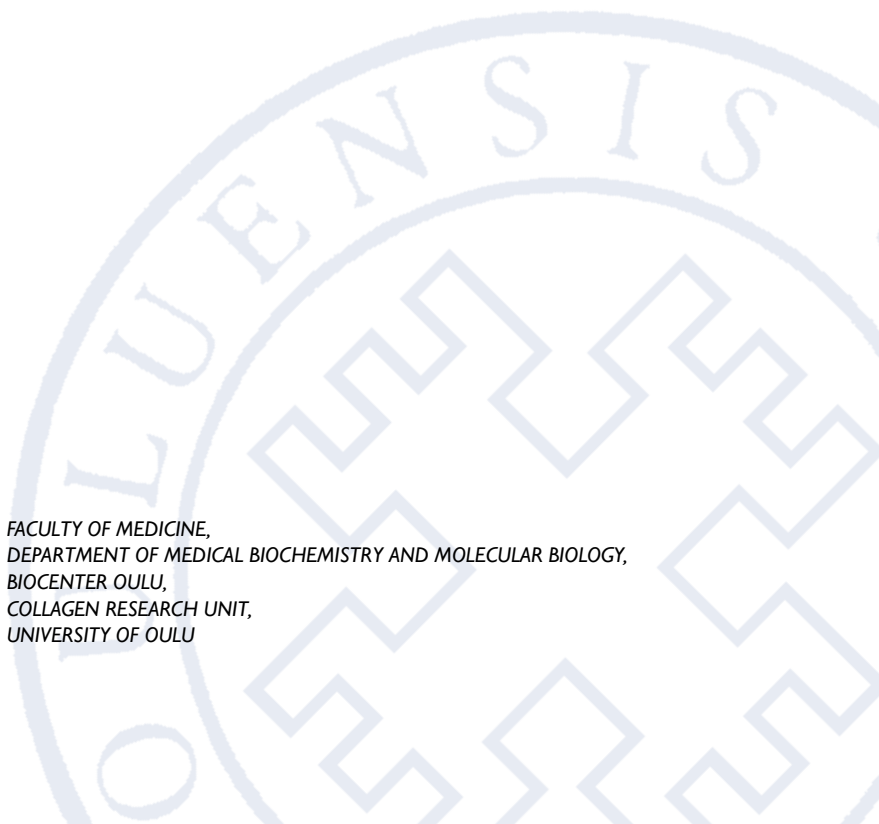
Katriina Keskiäho-Saukkonen

PROLYL 4-HYDROXYLASE

STUDIES ON COLLAGEN PROLYL 4-HYDROXYLASES
AND RELATED ENZYMES USING THE GREEN
ALGA CHLAMYDOMONAS REINHARDTII AND
TWO CAENORHABDITIS NEMATODE SPECIES
AS MODEL ORGANISMS

FACULTY OF MEDICINE,
DEPARTMENT OF MEDICAL BIOCHEMISTRY AND MOLECULAR BIOLOGY,
BIOCENTER OULU,
COLLAGEN RESEARCH UNIT,
UNIVERSITY OF OULU

D
MEDICA



KATRIINA KESKIAHO-SAUKKONEN

PROLYL 4-HYDROXYLASE

Studies on collagen prolyl 4-hydroxylases and related enzymes using the green alga *Chlamydomonas reinhardtii* and two *Caenorhabditis* nematode species as model organisms

Academic dissertation to be presented, with the assent of the Faculty of Medicine of the University of Oulu, for public defence in the Auditorium of the Medipolis Research Center (Kiviharjuntie 11), on May 25th, 2007, at 10 a.m.

Copyright © 2007
Acta Univ. Oul. D 930, 2007

Supervised by
Docent Johanna Myllyharju

Reviewed by
Associate Professor Simon Tuck
Professor Garry Wong

ISBN 978-951-42-8472-4 (Paperback)
ISBN 978-951-42-8473-1 (PDF)
<http://herkules.oulu.fi/isbn9789514284731/>
ISSN 0355-3221 (Printed)
ISSN 1796-2234 (Online)
<http://herkules.oulu.fi/issn03553221/>

Cover design
Raimo Ahonen

OULU UNIVERSITY PRESS
OULU 2007

Keskiaho-Saukkonen, Katriina, Prolyl 4-hydroxylase. Studies on collagen prolyl 4-hydroxylases and related enzymes using the green alga *Chlamydomonas reinhardtii* and two *Caenorhabditis* nematode species as model organisms

Faculty of Medicine, Department of Medical Biochemistry and Molecular Biology, Biocenter Oulu, Collagen Research Unit, University of Oulu, P.O. Box 5000, FI-90014 University of Oulu, Finland
Acta Univ. Oul. D 930, 2007
Oulu, Finland

Abstract

Collagen prolyl 4-hydroxylases (C-P4Hs) and related enzymes catalyze the hydroxylation of certain proline residues in animal collagens and plant hydroxyproline-rich proteins, respectively. Animal C-P4Hs and their isoenzymes have been characterized to date from humans, rodents, insects and nematodes. Most of the animal C-P4Hs are $\alpha_2\beta_2$ tetramers in which protein disulphide isomerase (PDI) serves as the β subunit, but the nematode C-P4Hs characterized so far have unique molecular compositions. Two P4Hs have been cloned from the plant *Arabidopsis thaliana* and one from the *Paramecium bursaria* *Chlorella* virus-1, these being monomeric enzymes.

This thesis reports on the identification of a large P4H family in the green alga *Chlamydomonas reinhardtii* and the cloning and characterization of one member, Cr-P4H-1. This is a soluble monomer that hydroxylates *in vitro* several peptides representing sequences found in *C. reinhardtii* cell wall proteins. Lack of its activity led to a defective cell wall structure, indicating that Cr-P4H-1 is essential for proper cell wall assembly and that the other P4Hs cannot compensate for the lack of its activity.

Two *C. elegans* genes, Y43F8B.4 and C14E2.4, predicted to code for C-P4H α subunit-like polypeptides were analyzed. Three transcripts were generated from Y43F8B.4, one of them coding for a functional C-P4H α subunit named PHY-4.1. C14E2.4 turned out not to be a C-P4H α subunit gene, as a frame-shift led to the omission of codons for two catalytically critical residues. PHY-4.1 formed active tetramers and dimers with PDI-2 and had unique substrate requirements in that it hydroxylated certain other proline-rich sequences besides collagen-like peptides. Inactivation of the Y43F8B.4 gene led to no obvious morphological abnormalities. Spatial expression of the *phy-4.1* transcript and PHY-4.1 polypeptide was localized to the pharynx and the excretory duct. Taken together, these data indicate that PHY-4.1 is not involved in the hydroxylation of cuticular collagens but is likely to have other substrates *in vivo*.

Cloning and characterization of the PHY-1 and PHY-2 subunits from the closely related nematode *Caenorhabditis briggsae* revealed distinct differences in assembly properties between the *C. elegans* and *C. briggsae* PHY-2 subunits in spite of their high amino acid sequence identity. Genetic disruption of *C. briggsae phy-1* resulted in a less severe phenotype than that observed in *C. elegans*, evidently on account of its more efficient assembly of the *C. briggsae* PHY-2 to an active C-P4H explaining the milder phenotype. Rescue of *C. elegans* and *C. briggsae phy-1* mutants was achieved by injection of a wild-type *phy-1* gene from either species.

Keywords: *Caenorhabditis briggsae*, *Caenorhabditis elegans*, *Chlamydomonas reinhardtii*, collagen, green alga, isoenzymes, nematode, prolyl 4-hydroxylase

Acknowledgements

This work was carried out at the Collagen Research Unit in the Department of Medical Biochemistry and Molecular Biology, University of Oulu, and at Biocenter Oulu.

I wish to express my deepest gratitude to Docent Johanna Myllyharju, who gave me the opportunity to prepare my thesis in her group. It has been a privilege to do so. I have been fortunate to have her as my supervisor, giving me guidance, support and encouragement during these years. Her contribution to the good team spirit has been invaluable. My sincere thanks go to Academy Professor Emeritus Kari Kivirikko for his endless optimism and inspiring ideas concerning the broad area of collagens and their enzymes, as well as preparing the ground for collagen research here in Oulu. Professor Taina Pihlajaniemi is acknowledged for providing excellent research facilities and an inspiring scientific atmosphere in the department and at Biocenter Oulu. I would also like to express my gratitude to Professor Peppi Karppinen, Head of the Department, for her positive attitude and for being such a pleasant colleague. I similarly wish to express my respect to Professor Emeritus Ilmo Hassinen, Professor Leena Ala-Kokko and Professor Seppo Vainio for their outstanding scientific work.

I am very grateful to Professor Garry Wong and Associate Professor Simon Tuck for their careful review of this thesis and valuable comments on it. Malcolm Hicks is acknowledged for his speedy revision of the language.

I am very grateful to Dr. Antony Page and Dr. Alan Winter for their invaluable collaboration and skilful guidance in *C. elegans* projects and for giving me the opportunity to visit their lab in Glasgow, Scotland. Raija Sormunen, as one of my co-authors, deserves sincere thanks for her expertise in the field of electron microscopy. Jussi Vuoristo, Ritva Nissi, Päivi Riihimaa, Gillian McCormack and M-A Felix are acknowledged for their contributions to the *C. elegans* and *C. briggsae* studies.

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. Risto Helminen is especially acknowledged for his kind, fast and professional computer support. Especial thanks go to Liisa Äijälä for her excellent, skilful technical assistance.

My warmest thanks go to the three wonderful women in our office, Antje, Maija and Päivi. It has been a privilege to share the office, and the cookies, with you. You have

made this feel like a joyful journey. Thank you for all the memorable moments, at work and outside. I will cherish them all. I would like to thank Reija Hieta especially warmly for her helpful advice on computer matters and for her cooperation in the “Chlamy” project, and also Liisa Kukkola for her positive attitude towards life and for sharing the *C. elegans* “fai-fai” project with me. I also wish to thank Kati Takaluoma for sharing the important day and dissertation party with me. Likewise I would like to thank all our group members, past and present, for creating the good spirit that prevails in our group.

I owe my deepest gratitude to my parents, Kaarina and Kauko, for their support and encouragement throughout my life, and my parents-in-law, Anna-Liisa and Juhani, are also warmly thanked. The “regular customers” of Club 6th Floor; lil'brou Jarkko, brou Jukka and Heidi deserve big hugs for sharing relaxing moments with me, I love you all and one sentence for you: “Hei-hei-hei, mikä tää on?”. The Kerhoto members are also acknowledged, especially Killi, Jori and Maija, Sussu and Petteri and Musse. Thank you for all the good moments. Let's have more of them. And finally, and above all, I owe my deepest love and gratitude to my husband Hannu Saukkonen and to our wonderful daughter Matilda. I thank Hannu for his never-ending encouragement, love and all the hugs that I got even without reason. You and Matilda are my driving force and there are no words to express my love for you.

This work was supported by the Finish Centre of Excellence Programme 2000-2005 of the Academy of Finland, the Biocenter Oulu and Pohjois-Suomen Kulttuurirahasto.

Oulu, May 2007

Katriina Keskiäho-Saukkonen

Abbreviations

AGP	arabinogalactan protein
Bli	blister phenotype
<i>B. mori</i>	<i>Bombyx mori</i>
bp(s)	base pair(s)
Cb	<i>Caenorhabditis briggsae</i>
<i>C. briggsae</i>	<i>Caenorhabditis briggsae</i>
cDNA	complementary DNA
Ce	<i>Caenorhabditis elegans</i>
<i>C. elegans</i>	<i>Caenorhabditis elegans</i>
C-P4H	collagen prolyl 4-hydroxylase
CSM	collagen-related structural motif
<i>D. melanogaster</i>	<i>Drosophila melanogaster</i>
Dpy	dumpy (short and fat) phenotype
dsRNA	double-stranded RNA
<i>E. coli</i>	<i>Escherichia coli</i>
ER	endoplasmic reticulum
GFP	green fluorescent protein
HIF	hypoxia-inducible factor
HIF-P4H	HIF prolyl 4-hydroxylase
HRGP	hydroxyproline-rich glycoprotein
K_i	inhibitory constant
K_m	Michaelis-Menten constant
Lon	long phenotype
mRNA	messenger RNA
P4H	prolyl 4-hydroxylase
PBCV-1	<i>Paramecium bursaria</i> <i>Chlorella</i> virus-1
PCR	polymerase chain reaction
PDI	protein disulphide isomerase
PHY	<i>C. elegans</i> C-P4H α subunit
PRP	proline-rich protein
PTL	potato tuber lectin
RACE	rapid amplification of cDNA ends

rbcS2	small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase
RNAi	RNA interference
Rol	roller phenotype
siRNA	small interfering RNA
Sma	small phenotype
Sqt	squat phenotype
TPR	tetratricopeptide repeat
V_{\max}	maximum reaction velocity
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 Keskiaho K, Hieta R, Sormunen R & Myllyharju J (2007) *Chlamydomonas reinhardtii* has multiple prolyl 4-hydroxylases, one of which is essential for proper cell wall assembly. *Plant Cell* 19: 256-269.
- II Winter AD, Keskiaho K, Kukkola L, McCormack G, Felix M-A, Myllyharju J & Page AP (2007) Differences in collagen prolyl 4-hydroxylase assembly between two *Caenorhabditis* nematode species despite high amino acid sequence identity of the enzyme subunits. *Matrix Biol*, 2007, Jan 30.
- III Keskiaho-Saukkonen K, Kukkola L, Page AP, Winter AD, Vuoristo J, Nissi R, Riihimaa P & Myllyharju J. Characterization of a novel *Caenorhabditis elegans* prolyl 4-hydroxylase with a unique substrate specificity and restricted expression in the pharynx and excretory duct. Manuscript.

Contents

Abstract	
Acknowledgements	
Abbreviations	
List of original articles	
Contents	
1 Introduction	13
2 Review of the literature	14
2.1 4-Hydroxyproline in animal and plant proteins	14
2.1.1 Vertebrate collagens	15
2.1.2 Invertebrate collagens	16
2.1.3 Plant proteins containing 4-hydroxyproline	20
2.2 Collagen prolyl 4-hydroxylases	22
2.2.1 Vertebrate collagen prolyl 4-hydroxylases	22
2.2.1.1 Molecular properties	22
2.2.1.2 Catalytic properties	24
2.2.1.3 Reaction mechanism	26
2.2.1.4 Collagen prolyl 4-hydroxylase I knock-out mice	28
2.2.2 Invertebrate collagen prolyl 4-hydroxylases	28
2.2.2.1 Nematode collagen prolyl 4-hydroxylases	28
2.2.2.2 <i>Drosophila melanogaster</i> prolyl 4-hydroxylases	30
2.2.3 Plant and viral prolyl 4-hydroxylases	32
2.2.3.1 Higher plant and algal prolyl 4-hydroxylases	32
2.2.3.2 <i>Paramecium bursaria</i> Chlorella virus-1 prolyl 4-hydroxylase	33
2.3 <i>Caenorhabditis elegans</i> and <i>Chlamydomonas reinhardtii</i> as model organisms	33
2.3.1 <i>C. elegans</i>	33
2.3.2 <i>C. reinhardtii</i>	36
3 Outlines of the present research	38
4 Materials and methods	39
5 Results	41

5.1 Characterization of the <i>C. reinhardtii</i> prolyl 4-hydroxylase gene family and detailed analysis of one of the encoded polypeptides, Cr-P4H-1.....	41
5.1.1 Amino acid sequence of <i>C. reinhardtii</i> P4H-1.....	42
5.1.2 Expression of recombinant Cr-P4H-1A in insect cells.....	42
5.1.3 Expression of recombinant Cr-P4H-1A and Cr-P4H-1B in <i>E. coli</i>	43
5.1.4 Catalytic properties of the recombinant Cr-P4H-1A and Cr-P4H-1B.....	43
5.1.5 Analysis of the <i>in vivo</i> function of Cr-P4H-1 by RNAi.....	44
5.2 Cloning and characterization of a novel <i>C. elegans</i> prolyl 4-hydroxylase.....	46
5.2.1 Molecular cloning of the <i>C. elegans</i> phy-4.....	46
5.2.2 Expression of recombinant PHY-4.1 and PHY-4.2 in insect cells.....	47
5.2.3 Catalytic properties of the recombinant P4H tetramers and dimers containing PHY-4.1 as the catalytic subunit.....	47
5.2.4 Analysis of the expression of phy-4.....	48
5.2.5 Analysis of the <i>in vivo</i> function of phy-4 by RNAi.....	49
5.3 Analysis of <i>C. briggsae</i> cuticle collagen prolyl 4-hydroxylases and comparison with those of <i>C. elegans</i>	49
5.3.1 Cloning of <i>C. briggsae</i> PHY-1, PHY-2 and PDI-2.....	49
5.3.2 Rescue experiments on <i>C. elegans</i> and <i>C. briggsae</i> phy-1 mutants using wild-type phy-1 and phy-2 genes from both species.....	49
5.3.3 RNAi of <i>C. briggsae</i> phy-1, phy-2 and pdi-2.....	50
5.3.4 Analysis of the collagen prolyl 4-hydroxylase forms generated from <i>C. briggsae</i> PHY-1, PHY-2 and PDI-2 in insect cells and <i>in vivo</i>	50
5.3.5 Analysis of the regions of the <i>C. elegans</i> and <i>C. briggsae</i> PHY polypeptides that determine their assembly properties.....	51
6 Discussion.....	53
6.1 <i>C. reinhardtii</i> has multiple prolyl 4-hydroxylases, one of which is essential for proper cell wall assembly.....	53
6.2 A novel member of the <i>C. elegans</i> prolyl 4-hydroxylase family.....	56
6.3 Assembly of cuticle collagen prolyl 4-hydroxylases in <i>C. elegans</i> and <i>C. briggsae</i>	58
References	
Original articles	

1 Introduction

Collagens are the most abundant proteins of the extracellular matrix and have many important structural and functional roles in tissues. The stability of the collagen triple helix is dependent on the presence of 4-hydroxyproline residues, which are formed in a co-translational and post-translational hydroxylation reaction catalyzed by collagen prolyl 4-hydroxylases (C-P4Hs). Most animal C-P4Hs characterized so far are tetrameric $\alpha_2\beta_2$ enzymes formed from a highly insoluble α subunit and protein disulphide isomerase (PDI), which serves as the β subunit. In nematodes, however, these subunits have been shown to assemble into active C-P4Hs with unique molecular compositions. The discovery of the existence of several α subunit isoforms in different animal species has shown that C-P4Hs actually comprise enzyme families. The current number of known α subunit isoforms is three in the human, mouse and rat, 19 in the fruitfly *Drosophila melanogaster*, and three in the nematode *Caenorhabditis elegans*.

Plants do not contain any collagens, but they have hydroxyproline-rich glycoproteins that are needed for the structural integrity of their cell walls. Plants also seem to have large P4H families, that of *Arabidopsis thaliana* consisting of at least six members, two of which have been cloned and characterized. A collagen-related structural motif has also been identified in some proteins of bacteria, bacteriophages and viruses. So far, only one viral P4H has been identified and characterized. The P4Hs from plant and viral sources differ from the animal C-P4Hs in that they are monomers.

A monomeric soluble P4H has previously been partially purified from the green alga *Chlamydomonas reinhardtii*. One of the aims of the present study was to clone this *C. reinhardtii* P4H, express it as a recombinant protein for determination of its kinetic properties and study its *in vivo* roles by RNA interference (RNAi).

Another aim was to identify and characterize potential additional *C. elegans* C-P4H α subunit isoforms. A fourth α subunit, PHY-4.1, was cloned, the catalytic properties of active recombinant C-P4H tetramers and dimers assembled from PHY-4.1 and *C. elegans* PDI-2 were analyzed, and the spatial expression pattern of *phy-4.1* was determined.

The present thesis also describes cloning of the major cuticle C-P4H α subunits, PHY-1 and PHY-2, from *C. briggsae*, a close relative of *C. elegans*, and analysis of their assembly properties relative to those of *C. elegans*. The rescue of *C. elegans* and *C. briggsae phy-1* mutants with inter-species wild-type genes was also analyzed.

2 Review of the literature

2.1 4-Hydroxyproline in animal and plant proteins

The majority of 4-hydroxyproline in animal proteins is found in collagens (see 2.1.1 and 2.1.2) and in other proteins containing collagen-like sequences. 4-Hydroxyproline plays an important role in stabilizing the collagen triple helix under physiological conditions, since non-hydroxylated collagen polypeptide chains cannot assemble into triple-helical collagen molecules at human body temperature. Proline is frequently found in the X positions of the -Gly-X-Y- repeats of collagens and 4-hydroxyproline in the Y positions, the hydroxylation of the Y-position prolines being catalyzed by C-P4Hs. The polypeptide chains of the most abundant collagen in vertebrates, type I collagen, contain about 1000 amino acid residues, of which about 100 are 4-hydroxyproline. (for reviews, see Kivirikko & Pihlajaniemi 1998, Myllyharju & Kivirikko 2004)

Elastic fibres are found in the extracellular matrix of many tissues, where they form three-dimensional meshworks of fine fibrils that are considered to provide support with flexibility for tissues (for a review, see Cleary & Gibson 1996). Elastin, which also has repeating -Gly-X-Y- triplets with 4-hydroxyproline in the Y positions, is the major component of elastic fibres, but it does not form triple-helical structures. The 4-hydroxyproline content of elastin varies, usually being about 10-25 residues per 1000 amino acids (for a review, see Kivirikko & Pihlajaniemi 1998).

A novel critical role for 4-hydroxyproline has recently been identified in the hypoxia-inducible transcription factor (HIF), an $\alpha\beta$ dimer (Ivan *et al.* 2001, Jaakkola *et al.* 2001). Hydroxylation of HIF occurs in a -Leu-X-X-Leu-Ala-Pro- sequence of its α subunit, catalyzed by a family of HIF-P4Hs that are distinct from the C-P4Hs (Bruick & McKnight 2001, Epstein *et al.* 2001, Ivan *et al.* 2002). This oxygen-dependent modification of HIF acts as a cellular oxygen sensor, the lack of which triggers a metabolic response when the oxygen supply becomes limiting. HIF is not hydroxylated in hypoxia and escapes the 4-hydroxyproline-mediated targeting to rapid proteasomal degradation, becomes stabilized and upregulates the expression of numerous genes involved in a variety of cellular hypoxia responses, including angiogenesis, erythropoiesis, glucose uptake, energy metabolism, apoptosis and vasomotor function (for reviews, see Kaelin 2005, Schofield & Ratcliffe 2005). This regulatory pathway is

probably evolutionarily conserved, since homologues of HIF subunits and HIF-P4H genes are found in the genomes of *C. elegans* and *D. melanogaster* (Bruck & McKnight 2001, Epstein *et al.* 2001).

Hydroxyproline-rich glycoproteins (HRGPs) are the major proteinaceous components of plant cell walls and often become covalently cross-linked into large meshworks. They have been identified in monocots, dicots and green algae and are subdivided into four distinct groups: the arabinogalactan proteins (AGPs), extensins, hydroxyproline-rich lectins and proline-rich proteins (PRPs). (for reviews, see Cassab 1998, Kieliszewski and Shpak, 2001)

A collagen-related structural motif (CSM) that has been identified in some proteins of bacteria, bacteriophages and viruses (Rasmussen *et al.* 2003) is believed to have a role in the attachment, elongation and trimerization of a variety of surface proteins in bacteria and viruses. Although CSMs have the characteristic -Gly-X-Y- repeats, they differ from the human collagens in several aspects (Medveczky *et al.* 1993, Smith *et al.* 1998, Rasmussen *et al.* 2003). Their proline content is significantly lower, and instead of frequently having proline and 4-hydroxyproline in the X and Y positions, their X positions are occupied by either proline, alanine or serine and their Y positions contain a higher proportion of threonine and glutamine. It is generally believed that CSMs do not contain 4-hydroxyproline, and that alternative means for stabilization of the triple-helical structure are provided by threonine in the Y positions, through indirect hydrogen bonding or glycosylations (Rasmussen *et al.* 2003). Although 4-hydroxyproline has been widely believed to be absent from bacterial and viral proteins, a P4H has been identified and cloned from the eukaryotic algal *Paramecium bursaria* *Chlorella* virus-1 (PBCV-1) (Eriksson *et al.* 1999).

2.1.1 Vertebrate collagens

Collagens have a significant role in maintaining the structural integrity of various tissues and also have many other important biological functions. At the moment at least 28 collagen types involving 43 distinct polypeptide chains (α chains) are known in vertebrates and more than 20 additional proteins have collagen-like domains. (Myllyharju & Kivirikko 2004, Ricard-Blum & Ruggiero 2005, Veit *et al.* 2005)

The basic structure of a collagen molecule consists of three polypeptide chains, called α chains, that are composed of -Gly-X-Y- repeats. In some collagens the α chains are identical (homotrimers), while others contain two or three different α chains (heterotrimers). The left-handedly coiled α chains wrap around each other to form a right-handed superhelical structure. The packing of the α chains into the triple helix requires that every third amino acid should be glycine which is small enough to fit into the restricted space in the centre of the helix. Proline is frequently found in the X position and 4-hydroxyproline in the Y position. (for reviews, see Myllyharju & Kivirikko 2004, Ricard-Blum & Ruggiero 2005)

The presence of 4-hydroxyproline residues in the Y positions is required for the thermal stability of collagen (Berg & Prockop 1973). As shown in X-ray diffraction studies, 4-hydroxyproline residues stabilize collagen by forming water bridges within the

α chains, between them and between different collagen molecules (Bella *et al.* 1994, 1995). The contribution of these water bridges to molecule stability is still controversial, however, since the triple helix of a (Pro-4Hyp-Gly)₁₀ repeat is stable in an anhydrous environment and replacement of 4-hydroxyproline with an electronegative 4(R)-fluoroproline leads to increased thermal stability (Holmgren *et al.* 1999). It has therefore been suggested that the stability of collagen relies rather on stereoelectronic effects that fix the pyrrolidine ring pucker and thus preorganize all three main-chain torsion angles (Jenkins & Raines 2002).

The majority of vertebrate collagens form supramolecular assemblies such as fibrils and networks, and they can be divided into eight subfamilies based on such assemblies or other features: (1) fibril-forming collagens (types I, II, III, V, XI, XXIV, XXVII), (2) fibril-associated collagens with interrupted triple helices and related collagens (types IX, XII, XIV, XVI, XIX, XX, XXI, XXII, XXVI, XXVIII), (3) the family of collagens forming hexagonal networks (types VIII and X), (4) the family of type IV collagens that are the major components of basement membranes, (5) type VI collagen, forming beaded filaments, (6) type VII collagen, forming anchoring fibrils in basement membranes, (7) collagens with transmembrane domains (types XIII, XVII, XXIII, XXV), and (8) collagen types XV and XVIII. (Myllyharju & Kivirikko 2004, Veit *et al.* 2005)

2.1.2 Invertebrate collagens

The collagens in nematodes are involved in the formation of two distinct structures, the cuticle, also called the exoskeleton, and the basement membranes (Figure 1).

The cuticle is a flexible structure composed of up to six layers and with a thickness dependent on the developmental stage and species of the nematode (Figure 1). It is composed of highly cross-linked, soluble and insoluble structural proteins, namely collagens, cuticulins and other minor proteins, lipids and carbohydrates. The major functions of the exoskeleton are to act as an impermeable barrier to the environment, allow movement, determine the body structure and permit growth through larval moults (for reviews, see Johnstone 2000, Page 2001). The cuticle of *C. elegans* consists mainly of small collagen-like proteins encoded by a multigene family of approximately 175 members (for reviews, see Johnstone 2000, Page 2001, Page & Winter 2003, Myllyharju & Kivirikko 2004). The cuticle is shed four times during the development of *C. elegans* and the cuticle collagen genes are accordingly expressed in a distinct temporal fashion in the hypodermal cells (for reviews, see Johnstone 2000, Page 2001, Page & Winter 2003, Myllyharju & Kivirikko 2004).

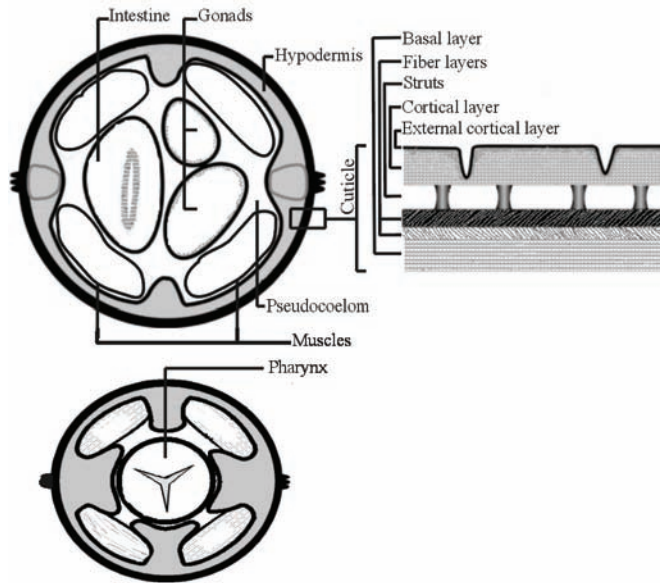


Fig. 1. Anatomy and cuticle structure of an adult *C. elegans* nematode. On the left, schematic cross-section through the central body (upper drawing) and head (lower drawing) regions. The locations of basement membranes inside the cuticle are drawn with black continuous lines. On the right, schematic structure of a longitudinal section of the cuticle. Modified from Kramer 1994 and Graham *et al.* 1997.

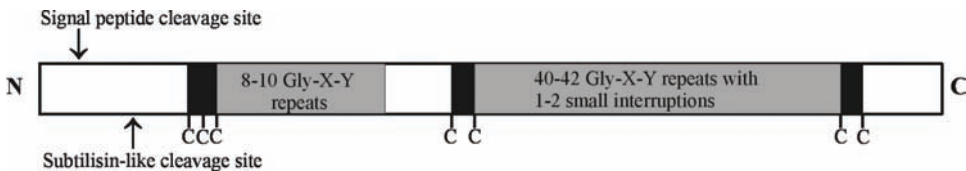


Fig. 2. Schematic representation of the structure of a *C. elegans* cuticle collagen.

The nematode cuticle collagen polypeptides are relatively small in size, 26-35 kDa (Figure 2) and have two collagenous domains, the N-terminal domain containing 8-10 -Gly-X-Y- repeats and the C-terminal one 40-42 -Gly-X-Y- repeats with 1-2 small interruptions. The collagenous domains are flanked by three clusters of cysteine residues. In addition to the collagenous domains the molecules have non-collagenous N and C-terminal regions. The N-terminal region is variable in size and contains homologous regions, including a signal peptide cleavage site and a proposed subtilisin-like cleavage site that is involved in the generation of the mature polypeptide from the procollagen precursor. The C-terminal region also varies in size, and cleavage has been shown to occur adjacent to a C-terminal tyrosine cross-linking site in the SQT-3 collagen. The processed cuticle collagens contain interchain disulphide bonds that can be either intramolecular or intermolecular, and they also contain tyrosine and putative γ carboxyl

glutamine-derived cross-links that are not found in vertebrate collagens. (Johnstone 2000, Page 2001, Page & Winter 2003, Myllyharju & Kivirikko 2004, Novelli *et al.* 2006)

Several mutations generated by random mutagenesis have been shown to affect *C. elegans* body morphology or viability. Seventeen cuticle collagen genes, named as *bli-1*, *bli-2*, *dpy-2*, *dpy-3*, *dpy-5*, *dpy-6*, *dpy-7*, *dpy-8*, *dpy-10*, *dpy-13*, *dpy-14*, *dpy-17*, *lon-3*, *rol-6*, *rol-8*, *sqt-1* and *sqt-3* (Kramer *et al.* 1988, von Mende *et al.* 1988, Kramer *et al.* 1990, Johnstone *et al.* 1992, Levy *et al.* 1993, van der Keyl *et al.* 1994, Bergmann *et al.* 1998, Johnstone 2000, Page 2001, Nyström *et al.* 2002, Page & Winter 2003, Myllyharju & Kivirikko 2004, Gallo *et al.* 2006, Thacker *et al.* 2006), and three basement membrane collagen genes *emb-9*, *let-2* and *cle-1* (Guo *et al.* 1989, 1991, Sibley *et al.* 1993, Ackley *et al.* 2001) have been identified by characterizing such mutations. Several *C. elegans* genes encoding collagen modifying enzymes (Figure 3) have been characterized so far, *phy-1*, *phy-2*, *phy-3* and *pdi-2* coding for C-P4H subunits, *let-268* for lysyl hydroxylase, 18 distinct *cyp* genes and 8 *fkf* genes for PPIases, *bli-4* and *dpy-31* for N and C-terminal proteinases, respectively, TPST-A for tyrosine O-sulphation, *dpy-11* for a thioredoxin, *pdi-3* for a transglutaminase, and *bli-3* and F53G12.3 for dual oxidases (Veijola *et al.* 1994, Thacker *et al.* 1995, Veijola *et al.* 1996, Friedman *et al.* 2000, Hill *et al.* 2000, Norman & Moerman 2000, Winter & Page 2000, Edens *et al.* 2001, Ko & Chow 2002, Riihimaa *et al.* 2002, Eschenlauer & Page 2003, Simmer *et al.* 2003, Novelli *et al.* 2004, Kim *et al.* 2005, Bell *et al.* 2006, Winter *et al.* 2007).

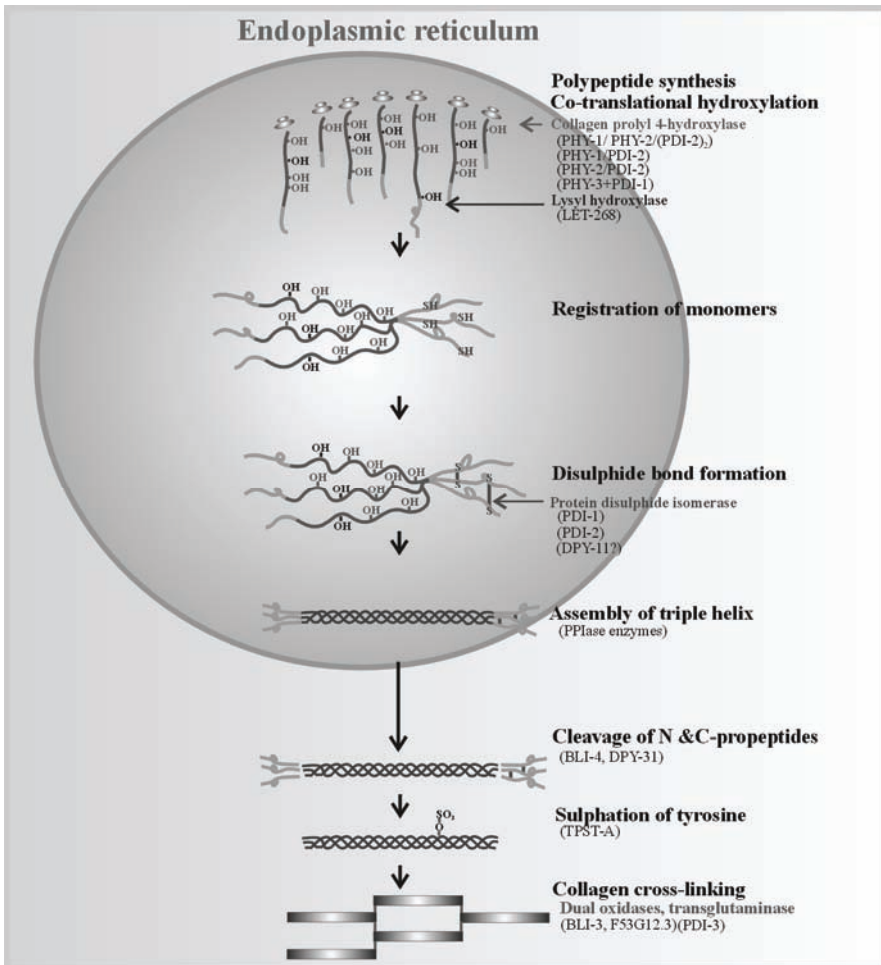


Fig. 3. The main steps in *C. elegans* collagen synthesis and the main enzymes involved. Modified from Myllyharju & Kivirikko 2004.

More recently, genome-wide and selective RNA interference (RNAi) screens have confirmed the importance of these previously identified components of cuticle and basement membrane synthesis and assembly, and more importantly, have identified novel genes that are involved in these aspects (Kamath *et al.* 2003, Simmer *et al.* 2003, Thein *et al.* 2003, Frand *et al.* 2005).

Nematode phenotypes associated with cuticle collagen gene mutations are divided into five basic types, described as (1) blister (*bli*), (2) dumpy (*dpy*), (3) roller (*rol*), (4) long (*lon*) and (5) squat (*sqt*). Nematodes with the blister phenotype have fluid-filled cuticle blisters that result in swelling and blistering of the cuticular material on from the surface of the animal. In the dumpy and long phenotypes the body length is shorter and longer, respectively, than in the wild type, whereas the roller phenotype results from a helical twisting of the body along its longitudinal axis. The phenotypes are not necessarily

exclusive, since different mutant alleles of a given gene can give rise to different phenotypes and some alleles show stage-specificity in the resulting phenotypes. The squat phenotype, for example, depends on the nature of the mutation, so that a heterozygous mutation evokes a roller appearance while a homozygous one generates a dumpy phenotype. (for reviews, see Johnstone 2000, Page 2001, Page & Winter 2003, Myllyharju & Kivirikko 2004)

Basement membranes are present in the pseudocoelomic face of the hypodermis and around the body wall muscles, pharynx, intestine and gonad in nematodes (Figure 1), and the basement membrane collagens of *C. elegans* are homologues of vertebrate type IV and XVIII collagens. The *C. elegans* type IV collagen homologue is composed of two different polypeptide chains encoded by the *emb-9* and *let-2* genes, while the type XVIII homologue is encoded by the *cle-1* gene. Mutations characterized in the *emb-9* and *let-2* genes cause embryonically lethal phenotypes (Gupta *et al.* 1997). The EMB-9 and LET-2 polypeptides are not expressed in the intestine, pharynx or hypodermis of *C. elegans*, but are secreted from the body wall muscles and delivered to these tissues (Graham *et al.* 1997, Gupta *et al.* 1997). Mutations in the CLE-1 protein, which is present in the basement membranes surrounding the gonad, pharynx, intestine and under-body wall muscles but accumulates to its highest levels in the nervous system, have been shown to affect cell migration and axon guidance and the organization and function of neuromuscular junctions (Ackley *et al.* 2001, 2003).

Fibrillar collagens have not been identified in nematodes nor in *D. melanogaster* or other insects (Lunstrum *et al.* 1988, Hynes & Zhao 2000). The *D. melanogaster* genome contains three conserved genes encoding basement membrane collagens, two α chains of a type IV collagen homologue and one of a type α XV/XVIII homologue (Hynes & Zhao 2000). Pericardin, which is synthesized in the pericardial cells and is involved in the morphogenesis and maintenance of the heart epithelium during dorsal ectoderm closure, also shows some homology with the *D. melanogaster* type IV collagen (Chartier *et al.* 2002), and several other *D. melanogaster* genes contain short sequences encoding collagen-like segments.

A basement membrane collagen has also been characterized from the silkworm *Bombyx mori* and shown to have roles in the formation of basement membranes, the encapsulation of foreign bodies during self-defence reactions and the metamorphic remodelling of the gut (Adachi *et al.* 2005). A second gene encoding a collagenous protein with multiple short collagenous domains separated by 9 interruptions has been identified in *B. mori* based on homology with known collagen sequences (Chareyre *et al.* 1996). The gene is expressed in the imaginal wing discs, epidermis and brain during the last larval moult and the spinning stage (Chareyre *et al.* 1996).

2.1.3 Plant proteins containing 4-hydroxyproline

The hydroxyproline-rich glycoproteins (HRGPs) of higher plants are usually divided into four groups. Arabinogalactan proteins (AGPs) are expressed on plant cell surfaces, being located on cell membranes, the cell wall and in gum exudates, and have important roles in plant growth, development and wound healing. AGPs are not covalently linked to the

cell wall and are therefore thought to have no structural function. Their protein content is less than 10%, and more than 90% of their mass consists of carbohydrates. The protein cores of AGPs are rich in 4-hydroxyproline, alanine, threonine, glycine and serine residues and contain short -Ala-4Hyp- repeats. The carbohydrate side chains are primarily *O*-linked to the hydroxyl groups of serine and 4-hydroxyproline. (for reviews, see Knox 1995, Cassab 1998, Showalter 2001)

Extensins are structural proteins of the cell walls of higher plants and are rich in 4-hydroxyproline and serine and varying combinations of valine, tyrosine, lysine and histidine. They characteristically possess a repeating -Ser-4Hyp-4Hyp-4Hyp-4Hyp-4Hyp-motif. Most of the 4-hydroxyproline residues are *O*-glycosylated with one to four arabinosyl residues, while many of the serine residues are glycosylated with a single galactose unit. In addition to serving as structural proteins, extensins are involved in development, wound healing and plant defence. (for reviews, see Showalter 1993, Cassab 1998)

The 4-hydroxyproline-rich lectins that comprise the third group of HRGPs are currently known only in solanaceous plants. They are localized in cell walls and are probably involved in various forms of cell-cell interaction through their carbohydrate binding properties. Potato tuber lectin (PTL), the most extensively studied 4-hydroxyproline-rich lectin, is composed of at least two distinct domains: one rich in serine and 4-hydroxyproline and containing a carbohydrate moiety, and the other rich in glycine and cysteine. The Ser-4Hyp-rich domain of PTL resembles the solanaceous extensins. (for a review, see Showalter 1993)

The proline-rich proteins (PRPs) found in plant cell walls and nodulins have important roles in various aspects of development and in nodule formation. PRPs contain essentially one repetitive motif, variations on lightly glycosylated -Pro-4Hyp-Val-Tyr-Lys- repeats, notably lacking serine. This repeat also occurs in extensins in combination with the -Ser-4Hyp-4Hyp-4Hyp-4Hyp-4Hyp- motif. (for a review, see Cassab 1998)

The cell walls of the volvocine algae, e.g. the multicellular *Volvox* and the unicellular *Chlamydomonas*, are assembled entirely from HRGPs. These HRGPs differ from those of higher plants, however. *C. reinhardtii* has at least two sets of HRGPs, the first of which is distributed in the multilayered cell walls of vegetative and gametic cells and the second is found in the cell walls of zygotic cells. The major HRGP constituents of the outer vegetative and gametic walls of *C. reinhardtii* are GP1, GP2 and GP3, which contain -Pro-Pro-Pro-, -Pro-X-Pro-, -Pro-X-X-Pro- and -Leu-Pro- sequence repeats, while zygotic cells contain a HRGP with -Pro-Pro-Pro- and -Ser-Pro- motifs (Adair & Apt 1990, Showalter 1993). According to amino acid analysis, the 4-hydroxyproline content of GP1 and GP2 proteins is around 32% and 14%, respectively (Adair & Apt 1990, Adair & Snell 1990, Ferris *et al.* 2001), and that of zygote cell wall proteins 22.5% (Woessner & Goodenough 1989). In addition, the plus and minus agglutinins of the algal cell that bring gametes of opposite mating types together during sexual reproduction are HRGPs that resemble the vegetative HRGPs (Ferris *et al.* 2005). Interestingly, two 4-hydroxyproline residues have also been detected in the L subunit of *C. reinhardtii* ribulose 1,5-bisphosphate carboxylase/oxygenase, which is not a HRGP (Taylor *et al.* 2001, Mizohata *et al.* 2002).

2.2 Collagen prolyl 4-hydroxylases

The C-P4Hs are located within the lumen of the endoplasmic reticulum (ER) and catalyze the formation of 4-hydroxyproline by the hydroxylation of proline residues in -X-Pro-Gly- repeats in collagens and many other proteins that have collagen-like domains (the -Gly-X-Y- repeats of collagens are given here as -X-Y-Gly- because of the hydroxylation properties of vertebrate C-P4Hs, see 2.2.1.2).

2.2.1 Vertebrate collagen prolyl 4-hydroxylases

2.2.1.1 Molecular properties

Three vertebrate C-P4H isoenzymes have been cloned to date from human, mouse and rat sources (Bassuk *et al.* 1989, Helaakoski *et al.* 1989, Hopkinson *et al.* 1994, Helaakoski *et al.* 1995, Annunen *et al.* 1997, Kukkola *et al.* 2003, Van Den Diepstraten *et al.* 2003), characterized and shown to be $\alpha_2\beta_2$ tetramers with a molecular weight of about 240 kDa. The active enzymes are composed of two distinct subunits, a catalytic α subunit (about 64 kDa) that differs between the isoenzymes and a β subunit (about 60 kDa) that is identical to PDI. The three α subunit isoforms, $\alpha(I)$, $\alpha(II)$ and $\alpha(III)$, associate with the same β subunit to form $[\alpha(I)]_2\beta_2$, $[\alpha(II)]_2\beta_2$ and $[\alpha(III)]_2\beta_2$ tetramers, called type I, II, and III C-P4Hs, respectively (for reviews, see Kivirikko & Myllyharju 1998, Kivirikko & Pihlajaniemi 1998, Myllyharju 2003, Myllyharju & Kivirikko 2004, Myllyharju 2005). Recombinant insect cell coexpression studies strongly argue against the formation of vertebrate C-P4H tetramers with dissimilar α subunits (Annuen *et al.* 1997).

The human $\alpha(I)$ subunit consists of 517 residues and a signal peptide of 17 additional residues, whereas the $\alpha(II)$ subunit consists of 514 amino acids and a signal peptide of 21 residues (Helaakoski *et al.* 1989, Annunen *et al.* 1997). The length of the human $\alpha(III)$ subunit is 525 amino acids plus a signal peptide of 19 amino acids (Kukkola *et al.* 2003, Van Den Diepstraten *et al.* 2003). The overall amino acid sequence identities of the human $\alpha(II)$ and $\alpha(III)$ subunits to the $\alpha(I)$ subunit are 64% and 35%, respectively, the highest degree of identity being observed in the C-terminal region, where catalytically important amino acids are located (Figure 4, see 2.2.1.2) (Annuen *et al.* 1997, Kukkola *et al.* 2003). The peptide-substrate-binding domain is separated from the catalytic domain (Figure 4) (Myllyharju & Kivirikko 1999, Hieta *et al.* 2003) and is located in the N-terminal part of the polypeptide. The structure of this domain has been determined at 2.3Å resolution in the case of the human $\alpha(I)$ subunit (see 2.2.1.2) (Pekkala *et al.* 2004).

The human $\alpha(I)$, $\alpha(II)$ and $\alpha(III)$ subunits contain five cysteines in conserved positions, the $\alpha(II)$ and $\alpha(III)$ subunits each having an additional non-conserved cysteine (Annuen *et al.* 1997, Kukkola *et al.* 2003). Site-directed mutagenesis studies on the $\alpha(I)$ subunit have indicated that intrachain disulphide bonds that are essential for the assembly of the $\alpha_2\beta_2$ tetramer are formed between the second and third and the fourth and fifth conserved cysteines (Lamberg *et al.* 1995). C-P4H tetramers have no interchain disulphide bonds (Nietfeld *et al.* 1981). All three α subunits have two attachment sites for N-linked

oligosaccharides, but enzymatic removal of sugars or mutations in the glycosylation sites have no effect on the tetramer assembly or catalytic activity of the type I and III C-P4Hs (Lamberg *et al.* 1995, Kukkola *et al.* 2003).

The genes encoding the three human α subunits are quite similar in size. The α (I) subunit gene is located on chromosome 10 and consists of 16 exons, while the α (II) and α (III) subunit genes are located on chromosomes 5 and 11 and consist of 16 and 13 exons, respectively. They are also very similar in their exon-intron organization, except for differences in the number of 5' exons and small differences in the lengths of some exons. Exons 9 and 10 of the α (I) gene and exons 12a and 12b in the α (II) gene are subject to mutually exclusive alternative splicing, but there is no evidence of alternative splicing in the α (III) gene (Pajunen *et al.* 1989, Helaakoski *et al.* 1994, Nokelainen *et al.* 2001, Kukkola *et al.* 2003).

PDI, which acts as the β subunit in all forms of vertebrate C-P4Hs, is a multifunctional enzyme that catalyzes the formation, reduction and isomerization of disulphide bonds during protein folding in the ER and also functions as a β subunit in the microsomal triglyceride transferase dimer and as a chaperone that assists the folding of many newly translated polypeptides. PDI has been identified and cloned from a wide range of species, including fungi, plants and animals. (for reviews, see Kivirikko & Myllyharju 1998, Kivirikko & Pihlajaniemi 1998, Freedman *et al.* 2002, Myllyharju 2003, Wilkinson & Gilbert 2004, Ellgaard & Ruddock 2005, Myllyharju 2005)

PDI is organized into four domains, *a*, *b*, *b'* and *a'*, and an acidic C-terminal extension *c*. The amino acid sequences of the *a* and *a'* domains are homologous with thioredoxin and each contains an independent -Cys-Gly-His-Cys- active site, while the catalytically inactive domains *b* and *b'* do not show sequence similarity to thioredoxin (for reviews, see Freedman *et al.* 2002, Wilkinson & Gilbert 2004, Ellgaard & Ruddock 2005). Nuclear magnetic resonance studies have shown, however, that all four domains have a thioredoxin fold (Kemink *et al.* 1996, 1997, 1999, Dijkstra *et al.* 1999). This was recently confirmed by crystallization of a full-length PDI from yeast (Tian *et al.* 2006). The domain *b'* provides the principal peptide-binding site of PDI and is also critical for the catalysis of isomerization but not oxidation reactions (Pirneskoski *et al.* 2004, Ellgaard & Ruddock 2005). Point mutations at this site do not inhibit C-P4H tetramer assembly, however (Koivunen *et al.* 2005). The minimum requirement for the assembly of an active C-P4H tetramer is fulfilled by the PDI domains *b'* and *a'*, while the domains *a* and *b* greatly enhance assembly but can in part be replaced by the corresponding domains of ERp57, a close homologue of PDI that is involved in the folding of newly synthesized glycoproteins (Koivunen *et al.* 1999, Pirneskoski *et al.* 2001). Recent studies have shown that binding sites in three PDI domains *a*, *b'*, and *a'*, contribute to efficient C-P4H assembly (Koivunen *et al.* 2005).

The active sites in the catalytic *a* and *a'* domains of PDI are not needed for the assembly or activity of the C-P4H tetramer (Vuori *et al.* 1992b). Deletion studies have shown that deletion of the PDI domain *c*, which contains the -Lys-Asp-Glu-Leu ER retention motif, leads to secretion of considerable amounts of both the free PDI polypeptide and an active recombinant C-P4H tetramer. Thus one of the functions of PDI as the β subunit of C-P4H is to mediate the retention of the enzyme tetramer within the lumen of the ER (Vuori *et al.* 1992b). When the C-P4H tetramer is dissociated or the α subunit is expressed in insect cells alone without the PDI polypeptide, the α subunits

form insoluble, inactive aggregates (Vuori *et al.* 1992a). Thus another important role for PDI as the C-P4H β subunit is to keep the α subunits in a non-aggregated, catalytically active conformation.

2.2.1.2 Catalytic properties

The C-P4Hs belong to the group of Fe(II)/2-oxoglutarate-dependent dioxygenases and catalyze the hydroxylation of peptide-linked prolines to 4-hydroxyproline. The reaction requires Fe^{2+} , 2-oxoglutarate, molecular oxygen and ascorbate as cosubstrates (for reviews, see Kivirikko & Myllyharju 1998, Kivirikko & Pihlajaniemi 1998, Myllyharju 2003, 2005). A catalytic site for the C-P4H α subunits comprises a set of separate locations for binding of the cosubstrates and is distinct from the peptide-substrate-binding domain (for reviews, see Myllyharju 2003, 2005) (Figure 4). Site-directed mutagenesis studies have indicated that three conserved residues, His412, Asp414 and His483, act as the Fe^{2+} binding ligands in the human $\alpha(\text{I})$ subunit (Lamberg *et al.* 1995, Myllyharju & Kivirikko 1997) (Figure 4). Mutation of either His412 or His483 to serine completely inactivates the enzyme (Lamberg *et al.* 1995, Myllyharju & Kivirikko 1997), as does mutation of Asp414 to alanine or asparagines, whereas a glutamate in that position reduces the activity by $\sim 85\%$ (Myllyharju & Kivirikko 1997). The 2-oxoglutarate binding site is divided into three distinct subsites: subsite I is formed by the amino acid Lys493 in the human $\alpha(\text{I})$ subunit (Myllyharju & Kivirikko 1997), which binds the C5 carboxyl group of 2-oxoglutarate (Figure 4), while subsite II consists of two *cis*-positioned coordination sites of the enzyme-bound Fe^{2+} and is chelated by the C1-C2 moiety of 2-oxoglutarate and subsite III involves a hydrophobic binding site in the C3-C4 region of the cosubstrate (Hanuske-Abel & Günzler 1982, Kivirikko & Myllyharju 1998, Kivirikko & Pihlajaniemi 1998, Hanuske-Abel & Popowicz 2003, Myllyharju 2003, 2005). His501, an additional critical residue at the catalytic site of the human $\alpha(\text{I})$ subunit, is involved in both the binding of the C1 carboxyl group of 2-oxoglutarate to Fe^{2+} and the decarboxylation of this cosubstrate (Myllyharju & Kivirikko 1997) (Figure 4). The ascorbate binding site contains two *cis*-positioned coordination sites of the enzyme-bound iron and is thus partially identical to the binding site of 2-oxoglutarate, but does not become bound at subsite I of the 2-oxoglutarate-binding site. Molecular oxygen is thought to be bound to the Fe^{2+} end-on in an axial position, producing the dioxygen unit (Hanuske-Abel & Günzler 1982, Kivirikko & Myllyharju 1998, Kivirikko & Pihlajaniemi 1998, Hanuske-Abel & Popowicz 2003, Myllyharju 2003, 2005).



Fig. 4. Schematic representation of the human C-P4H $\alpha(\text{I})$ subunit.

The amino acid residues identified by site-directed mutagenesis of the human $\alpha(I)$ subunit as being critical for the binding of Fe^{2+} and 2-oxoglutarate (Figure 4) are conserved in all animal C-P4Hs and plant and viral P4Hs studied so far, as well as in other 2-oxoglutarate dependent dioxygenases (see below), with the exception that in the latter the basic residue binding the C5 carboxyl group of 2-oxoglutarate is mostly an arginine instead of a lysine. The two Fe^{2+} binding histidines are located about 50-70 residues apart, while the aspartate is always located in position +2 with respect to the first Fe^{2+} binding histidine. The lysine or arginine that binds the C-5 carboxyl group of 2-oxoglutarate is located in position +9 or +10 with respect to the second Fe^{2+} binding histidine, the only exception being the HIF asparaginyl hydroxylase, where this residue is located between the Fe^{2+} binding histidines. (Kivirikko & Myllyharju 1998, Kivirikko & Pihlajaniemi 1998, Aravind & Koonin 2001, Myllyharju 2003, 2005, Clifton *et al.* 2006)

The C-P4Hs hydroxylate proline residues only in peptide linkages, and do not act on the corresponding free amino acids. A tripeptide X-Pro-Gly fulfils the minimum substrate sequence requirement for C-P4Hs, whereas Gly-X-Pro, Gly-Pro-Ala or Pro-Gly-X do not become hydroxylated. Therefore, prolines in the Y position preceding the glycine of -X-Y-Gly- sequences in collagens are hydroxylated. A C-P4H from the cuticular epithelium of earthworms constitutes an exception to this rule, in that a (Pro-Ala-Gly)_n is a better substrate than a peptide in which proline precedes a glycine. The rate of hydroxylation by vertebrate C-P4Hs is affected by the amino acid present in the X position, as proline in this position results in a high maximal reaction velocity (V_{\max}) while alanine, leucine, arginine, valine and glutamate give a lower V_{\max} . The presence of glycine or sarcosine in the X position of the (X-Pro-Gly)_n or a triple-helical conformation of the collagenous peptide completely prevents hydroxylation (for reviews, see Kivirikko *et al.* 1992, Kivirikko & Pihlajaniemi 1998). The chain length of the peptide substrate influences its K_m value for vertebrate C-P4Hs, in that the value decreases with increasing chain length, while the V_{\max} is not affected (for reviews, see Kivirikko *et al.* 1992, Kivirikko & Myllyharju 1998, Kivirikko & Pihlajaniemi 1998). This finding has been explained by a processive mechanism of binding of long peptide substrates in which an enzyme-substrate complex is formed at one catalytic site of the tetramer upon the first meeting and the other catalytic site then has a high chance of facing another hydroxylatable region of the peptide substrate. This mechanism requires the presence of two active catalytic sites and would lead to a low K_m for a long peptide by overcoming diffusional constraints on the association between the enzyme and the various substrate sites present in the peptide (de Waal & de Jong 1988, de Jong *et al.* 1991).

Studies with individual recombinant peptide-substrate-binding domains of the human type I and type II C-P4Hs have shown that they bind longer peptides more effectively than shorter ones, this most probably being the major determinant for the low K_m values of C-P4Hs for long peptide substrates (Hieta *et al.* 2003). This is supported by the fact that an active *C. elegans* C-P4H dimer that contains only one catalytic subunit has a similar preference for long substrates (Kukkola *et al.* 2004). These data indicate that the more efficient hydroxylation of long peptide substrates results from their higher binding affinity for the peptide-substrate-binding domain rather than from the proposed processive mechanism. Nuclear magnetic resonance characterization and crystallization of the peptide-substrate-binding domain of the human $\alpha(I)$ subunit at a resolution of 2.3 Å have shown that it is a helical protein consisting of five antiparallel α -helices and a loop

that is located between helices $\alpha 2$ and $\alpha 3$. It was found to belong to the family of tetratricopeptide repeat (TPR) domains that are involved in many protein-protein interactions. The most prominent surface feature of the domain is a deep groove on its concave side that is lined with tyrosine residues. Site-directed mutagenesis studies have indicated that these tyrosines are important for peptide binding. Previous studies of the conformational features of synthetic peptide substrates required for their hydroxylation by a C-P4H tetramer have suggested that a minimal requirement for proline 4-hydroxylation may be a sequence in a poly(L-proline) type II helix conformation followed by a β -turn in the Pro-Gly segment. This helix conformation may be necessary for effective interaction at the peptide-substrate-binding domain, whereas the β -turn may be essential for hydroxylation at the catalytic site (Hieta *et al.* 2003, Pekkala *et al.* 2004).

The crystal structures of several members of the Fe(II)/2-oxoglutarate-dependent dioxygenases have been resolved (for a review, see Clifton *et al.* 2006), including deacetoxycephalosporin C synthase (Valegård *et al.* 1998), clavaminic acid synthase (Zhang *et al.* 2000), anthocyanidin synthase (Wilmouth *et al.* 2002), the taurine/ α -ketoglutarate dioxygenases (Elkins *et al.* 2002), HIF asparaginyl hydroxylase (Dann *et al.* 2002, Elkins *et al.* 2003, Lee *et al.* 2003) and a HIF-P4H (McDonough *et al.* 2006). These all possess a β -strand jellyroll core fold. In addition, a related Fe(II)-dependent dioxygenase, isopenicillin N synthase, that does not utilize 2-oxoglutarate as a cosubstrate has a similar core fold (Roach *et al.* 1995). The jellyroll fold consists of 8 β -strands that form two 4-stranded sides. The additional α -helices, β -strands and loops that surround this core contribute to protein stability (for reviews, see Hausinger 2004, Clifton *et al.* 2006). All attempts to crystallize a vertebrate C-P4H $\alpha_2\beta_2$ tetramer have so far been unsuccessful. Structural studies of the isolated full-length α subunit are not possible as it aggregates in the absence of the PDI/ β subunit. In view of the known structures of Fe(II)/2-oxoglutarate-dependent dioxygenases, it is highly likely that the catalytic sites of the C-P4H α subunits also have the jellyroll core fold. HIF-P4Hs differ distinctly from C-P4Hs in that they hydroxylate a single proline in a -Leu-X-X-Leu-Ala-Pro- sequence and not in -X-Pro-Gly- repeats. Furthermore, they are homodimers and not $\alpha_2\beta_2$ tetramers like the vertebrate C-P4Hs. Therefore, the structure of a C-P4H tetramer would reveal important data on how the α and β subunits assemble into the tetramer, how the peptide-substrate-binding domains and the catalytic sites are arranged in the α subunits, and how the collagen polypeptide chain bound to the peptide-substrate-binding domain is presented to the catalytic site.

2.2.1.3 Reaction mechanism

The hydroxylation reaction catalyzed by C-P4Hs requires Fe^{2+} , 2-oxoglutarate, O_2 and ascorbate as cosubstrates and involves an oxidative decarboxylation of 2-oxoglutarate. Kinetic studies and mechanistic studies based on the crystal structures of the Fe(II)/2-oxoglutarate-dependent dioxygenases indicate that the hydroxylation reaction involves an ordered binding of Fe^{2+} , 2-oxoglutarate, substrate and molecular oxygen to the enzyme and a release of the reaction products in the reverse order, except that Fe^{2+} is not released

between catalytic cycles (for reviews, see Kivirikko & Myllyharju 1998, Kivirikko & Pihlajaniemi 1998, Myllyharju 2003, Schofield & Ratcliffe 2005).

2-Oxoglutarate is a highly specific requirement for the hydroxylation reaction and during the first half of the reaction it is stoichiometrically decarboxylated to succinate in such a way that one atom of the dioxygen molecule is consumed. Decarboxylation leads to the formation of a highly reactive iron-oxo complex, a ferryl ion, which is the active intermediate in the transfer of oxygen to proline. In the second half of the reaction the ferryl ion hydroxylates the proline residue in the peptide substrate so that the other atom of the dioxygen molecule is incorporated into the forming hydroxyl group. This restores the Fe^{2+} at the catalytic site and completes the reaction cycle. (Hanauske-Abel & Günzler 1982, Kivirikko *et al.* 1992, Kivirikko & Myllyharju 1998, Kivirikko & Pihlajaniemi 1998, Myllyharju 2003, 2005)

Ascorbate is not needed for the hydroxylation reaction, but it is consumed stoichiometrically in an uncoupled decarboxylation of 2-oxoglutarate, i.e. decarboxylation without subsequent hydroxylation of the peptide substrate. The reactive ferryl ion is converted to Fe^{3+}O^- in uncoupled reactions and the next catalytic cycle can progress only when this Fe^{3+}O^- has been reduced by ascorbate. Since the C-P4Hs catalyze such uncoupled decarboxylation cycles even in the presence of a saturating concentration of a peptide substrate, the role of ascorbate in the reaction is to serve as an alternative oxygen acceptor in the case of uncoupled reaction cycles. (Myllylä *et al.* 1984, Kivirikko *et al.* 1992, Kivirikko & Myllyharju 1998, Kivirikko & Pihlajaniemi 1998, Myllyharju 2003, 2005)

The vertebrate C-P4H isoenzymes have quite similar K_m values for their cosubstrates, but differences exist in their K_m values for peptide substrates and K_i values for peptide inhibitors. The K_m values of the human type I and type III C-P4Hs for a collagen-like peptide substrate (Pro-Pro-Gly)₁₀ are 15 and 24 μM , respectively, while that of the type II enzyme is higher, 95 μM . Poly(L-proline) is an efficient competitive inhibitor of type I C-P4H, the IC_{50} value for poly(L-proline) M_r 5000-7000 being only 6 μM , but it is a weak inhibitor of the type II enzyme, with a 50-fold higher IC_{50} . The human type III C-P4H is intermediate between the other two C-P4Hs in this respect, with an IC_{50} of 30 μM (Annunen *et al.* 1997, Kukkola *et al.* 2003). Site-directed mutagenesis studies have shown that these properties reflect differences in the amino acid sequences of the peptide-substrate-binding domains of the human $\alpha(\text{I})$ and $\alpha(\text{II})$ subunits (see above), the presence of a glutamate and a glutamine in the $\alpha(\text{II})$ subunit in positions corresponding to Ile182 and Tyr233 in the $\alpha(\text{I})$ subunit explaining most of the lack of poly(L-proline) binding to the type II enzyme (Myllyharju & Kivirikko 1999).

Several compounds inhibit C-P4Hs competitively with respect to some of its cosubstrates or the peptide substrate (e.g. poly(L-proline), see above). Divalent cations such as Zn^{2+} are competitive inhibitors with respect to Fe^{2+} , the K_i value of human type I C-P4H for Zn^{2+} being about 1 μM . Competitive inhibitors with respect to 2-oxoglutarate have functional groups that can bind to the ferrous ion at subsite II and also interact with the other subsites of the 2-oxoglutarate binding site (for reviews, see Kivirikko *et al.* 1992, Kivirikko & Myllyharju 1998, Kivirikko & Pihlajaniemi 1998, Myllyharju 2003, 2005). Two examples of such inhibitors are pyridine 2,4-dicarboxylate and pyridine 2,5-dicarboxylate, which have IC_{50} values of 8 to 11 μM and 1 to 7 μM for the three human C-P4Hs, respectively (Kukkola *et al.* 2003).

2.2.1.4 *Collagen prolyl 4-hydroxylase I knock-out mice*

Expression studies have shown that type I C-P4H is the main form in most cell types and tissues, the type II enzyme is the predominant form in chondrocytes, osteoblasts, endothelial cells and cells in epithelial structures, whereas the type III enzyme is expressed at low levels in many fetal and adult tissues and in many cell types (Annunen *et al.* 1998, Nissi *et al.* 2001, Kukkola *et al.* 2003). It has been shown recently that mice lacking a functional C-P4H α (I) subunit gene die at the embryonic stage, between E10.5 and E11.5, and show a general developmental delay, abnormal basement membranes and frequently ruptured capillary walls (Holster *et al.* 2007). The basement membranes essentially lack type IV collagen, the amount of soluble collagen IV being increased in the null embryos, indicating that the collagen IV molecules are incapable of assembling into insoluble supramolecular network structures. Surprisingly, type I and type III collagen fibrils appeared morphologically normal in the null embryos, the only apparent deviation being a slight increase in their diameters, the reason for which is still somewhat unclear. The null mice exhibit about a 80% reduction in total C-P4H activity, the remaining activity being due to the other two isoenzymes (Holster *et al.* 2007).

2.2.2 *Invertebrate collagen prolyl 4-hydroxylases*

2.2.2.1 *Nematode collagen prolyl 4-hydroxylases*

Three distinct *C. elegans* genes coding for C-P4H α subunits referred to as PHY-1, PHY-2 and PHY-3 have been cloned and characterized to date (Veijola *et al.* 1994, Friedman *et al.* 2000, Hill *et al.* 2000, Winter & Page 2000, Riihimaa *et al.* 2002), and likewise three PDI isoforms have been characterized from *C. elegans*, termed PDI-1, PDI-2 and PDI-3 (Veijola *et al.* 1996, Natsuka *et al.* 2001, Eschenlauer & Page 2003). The assembly properties of recombinant *C. elegans* PHY-1 and PHY-2 polypeptides differ distinctly from those of the vertebrate C-P4H α subunits. The PHY-1 and PHY-2 polypeptides combine with PDI-2 to form a unique mixed tetramer PHY-1/PHY-2/(PDI-2)₂, while active PHY-1/PDI-2 and PHY-2/PDI-2 dimers can also be formed, although much less efficiently (Veijola *et al.* 1994, Myllyharju *et al.* 2002).

The assembly of the *C. elegans* mixed tetramer was found to be highly species-specific, as the PHY-1, PHY-2 and PDI-2 polypeptides cannot be replaced by human or *D. melanogaster* α and β subunits in recombinant coexpression studies (Myllyharju *et al.* 2002). Also, recombinant *C. elegans* PDI-1 cannot replace PDI-2 as the β subunit in the mixed tetramer or the PHY-1/PDI-2 and PHY-2/PDI-2 dimers (Veijola *et al.* 1996, Myllyharju *et al.* 2002). Interestingly, human PDI can assemble into an active dimer with *C. elegans* PHY-1, and actually much more efficiently than the *C. elegans* PDI-2, although it is incapable of acting as the β subunit in the mixed tetramer (Veijola *et al.* 1996, Myllyharju *et al.* 2002). Recombinant expression studies with *C. elegans* PHY-3 have shown that C-P4H activity is generated when it is coexpressed with PDI-1, but not with PDI-2 (Riihimaa *et al.* 2002). It is not known at present, however, whether PHY-3 forms a dimer or a tetramer with PDI-1, or whether the role of PDI-1 is just to assist in

the folding of PHY-3 (Riihimaa *et al.* 2002). It is likely that the third *C. elegans* PDI isoform, PDI-3, does not function as a C-P4H β subunit, but that its role is related to other functions involved in ECM assembly, most probably cross-linking of collagens (Eschenlauer & Page 2003).

The *C. elegans* PHY-1 polypeptide consists of 543 amino acids and a signal peptide of 16 additional residues, while PHY-2 is slightly shorter, 523 amino acids and a 16-amino-acid signal peptide (Veijola *et al.* 1994, Friedmann *et al.* 2000, Winter & Page 2000). The processed PHY-1 and PHY-2 polypeptides are 57% identical to each other, the highest degree of identity being found around the catalytically critical residues, while a major difference exists in their extreme C-terminal regions (Friedman *et al.* 2000, Winter & Page 2000). The processed PHY-1 polypeptide has a C-terminal extension of 32 residues that is not found in any other animal C-P4H α subunit characterized so far. The role of this extension is still unclear and it has been demonstrated that it is not involved in determination of the formation of the mixed tetramer (Myllyharju *et al.* 2002). The amino acid sequence identities of PHY-1 and PHY-2 to the human α (I) and α (II) subunits are 43-46% (Friedman *et al.* 2000, Winter & Page 2000). The PHY-3 polypeptide is considerably shorter than the PHY-1, PHY-2 and vertebrate α subunits, the length of the processed PHY-3 being only 295 amino acids (Riihimaa *et al.* 2002). The identity of PHY-3 to the corresponding regions of PHY-1 and PHY-2 and the human α (I) and α (II) subunits is 23-30% (Riihimaa *et al.* 2002).

The catalytic properties of the recombinant *C. elegans* PHY-1/PHY-2/(PDI-2)₂ tetramer are very similar to those of human C-P4H tetramers. The K_m values for cosubstrates and the peptide substrate (Pro-Pro-Gly)₁₀ and K_i values for certain inhibitors are comparable to those of the human enzymes, with the exception that the K_m value of the *C. elegans* mixed tetramer for 2-oxoglutarate is slightly higher. The *C. elegans* mixed tetramer is not inhibited by poly(L-proline), and thus resembles the human type II C-P4H in this respect (Myllyharju *et al.* 2002).

Reporter gene expression studies (Hill *et al.* 2002, Winter & Page 2002) and immunolocalization studies (Myllyharju *et al.* 2002) have shown that the *C. elegans phy-1*, *phy-2* and *pdi-2* genes and their encoded polypeptides are coexpressed in hypodermal cells in a cyclical fashion that coincides with the cuticle moulting cycle, suggesting that they have a role in cuticle collagen synthesis. The *phy-1* gene has been shown to correspond to the *dpy-18* locus, mutations in which cause a dumpy (short and fat) phenotype related to the decreased 4-hydroxyproline content of the cuticle collagens and their concomitant inability to support normal body morphology (Friedman *et al.* 2000, Hill *et al.* 2000, Winter & Page 2000). On the other hand, *phy-2* null mutants have a wild-type phenotype, suggesting that *phy-2* is not necessary for proper cuticle synthesis (Friedman *et al.* 2000, Winter & Page 2000). Simultaneous inactivation of both *phy-1* and *phy-2* genes is nevertheless lethal to the embryo, which is also the outcome in *pdi-2* null mutants (Friedman *et al.* 2000, Winter & Page 2000). These mutant embryos develop normally until the first cuticle is required to maintain the elongated shape, after which they collapse, fail to hatch, and eventually die.

The above genetic observations can be explained by the assembly properties of the PHY-1, PHY-2 and PDI-2 polypeptides. Analysis of protein samples from wild-type, *dpy-18* and *phy-2* mutant nematodes has shown that the mixed tetramer is the major C-4H form in the wild type, while small amounts of PHY-1/PDI-2 dimers are also observed,

whereas PHY-2/PDI-2 dimers are not detected (Myllyharju *et al.* 2002). In *dpy-18* mutants that lack the PHY-1 polypeptide and concomitantly the mixed tetramer and the PHY-1/PDI dimer, assembly of PHY-2/PDI-2 dimers is increased (Myllyharju *et al.* 2002), but this is not sufficient to compensate fully for the lack of the other forms, as evidenced by the abnormal phenotype of the mutant nematodes and the marked reductions, more than 95% and 18-60%, in the total C-P4H activity and 4-hydroxyproline content of cuticle collagens, respectively (Winter & Page 2000, Myllyharju *et al.* 2002). The lack of the mixed tetramer in the *phy-2* null mutants is compensated for by a marked increase in the assembly of the PHY-1/PDI-2 dimer, resulting in 50-60% and 64% residual C-P4H activity and 4-hydroxyproline content of cuticle collagens, respectively, relative to the wild type, these levels obviously being sufficient to support normal cuticle synthesis in the *phy-2* mutants (Friedman *et al.* 2000, Winter & Page 2000, Myllyharju *et al.* 2002).

The *C. elegans phy-3* gene has a highly restricted expression pattern. It is expressed in the spermatheca, the specialized region of the gonad where the oocytes are fertilized, but this is restricted temporally to embryos and late larval and adult nematodes. Inactivation of the *phy-3* gene leads to no obvious morphological abnormalities in the nematodes, nor are any defects found in their fertility or behaviour. The only observed change was that the 4-hydroxyproline content of the early embryos was reduced by about 90%. It is thus probable that PHY-3 hydroxylates collagens in the early embryos, probably those of the egg shells (Riihimaa *et al.* 2002).

C-P4H α subunits have also been characterized from the filarial nematodes *Brugia malayi* (Winter *et al.* 2003) and *Onchocerca volvulus* (Merriweather *et al.* 2001). The *B. malayi* PHY-1 consists of 524 amino acids and a signal peptide of 17 additional residues and shows 59% and 53% identity to the *C. elegans* PHY-1 and PHY-2, respectively. It is highly unusual, however, in that the recombinant polypeptide assembles into an active, soluble C-P4H homotetramer in the absence of any PDI. The *B. malayi phy-1* mRNA is expressed in all stages analyzed, L1, L3, L4 and adult nematodes, with peaks in abundance corresponding to the moults. The native spatial localization of *B. malayi phy-1* is not yet known, but heterologous reporter gene studies in *C. elegans* indicate hypodermal expression and thus a role in cuticle collagen synthesis (Winter *et al.* 2003). The genome of *O. volvulus* encodes at least three polypeptides that are homologous to C-P4H α subunits and one of these has been cloned and characterized in detail. It consists of 549 amino acids and a signal peptide of 23 additional residues and has a similar C-terminal extension to *C. elegans* PHY-1, but its role is also unclear. *O. volvulus phy-1* is expressed in L3 and adult parasites and the recombinant polypeptide combines with *O. volvulus* PDI to form an active C-P4H (Merriweather *et al.* 2001).

2.2.2.2 *Drosophila melanogaster* prolyl 4-hydroxylases

The *D. melanogaster* genome contains altogether 19 C-P4H α subunit-related genes, of which seven have been cloned as full-length cDNAs and one characterized at the protein level (Annunen *et al.* 1999, Abrams & Andrew 2002). The lengths of the encoded polypeptides range from 481 to 550 amino acids. Six of the 10 genes analyzed show a

tissue-specific embryonic expression pattern: two in the salivary gland, two in mouthpart precursors, one in the proventriculus and one in the epidermis (Abrams & Andrew 2002). The relatively large number of C-P4H α subunit-related genes in *D. melanogaster* is highly unexpected, since its genome contains only three genes coding for type IV and XVIII collagen polypeptide chains. It has therefore been suggested that some of the *D. melanogaster* C-P4Hs hydroxylate proline residues in proteins other than collagens (Abrams & Andrew 2002, Myllyharju & Kivirikko 2004).

The first α subunit to be cloned from *D. melanogaster*, PH4 α MP, assembles into an active tetramer with *D. melanogaster* PDI when these are coexpressed as recombinant polypeptides, and shows 30-35% identity to the human α (I) and α (II) subunits and the *C. elegans* PHY-1 polypeptide at the amino acid level (Annunen *et al.* 1999). Some distinct differences are found between the catalytic properties of the recombinant *D. melanogaster* C-P4H and those of human type I and II C-P4Hs, these differences being reflected in the amino acid sequences of the catalytic and peptide-substrate-binding domains of the α subunits. The fifth critical residue corresponding to the His501 in the human α (I) subunit, which probably assists binding of the C-1 carboxyl group of 2-oxoglutarate to the Fe²⁺ atom (Myllyharju & Kivirikko 1997), is replaced by an arginine in the *D. melanogaster* PH4 α MP (Annunen *et al.* 1999). The recombinant *D. melanogaster* enzyme had a higher K_m for 2-oxoglutarate, a lower reaction velocity and a higher percentage of uncoupled decarboxylation than the human enzymes, but although mutation of the PH4 α MP arginine to histidine reduced the percentage of uncoupled decarboxylation, it did not increase the reaction velocity or reduce the K_m for 2-oxoglutarate. The amino acid sequence of the PH4 α MP peptide-substrate-binding domain shows a relatively low identity to those of the human α (I) and α (II) subunits, and a collagen-like peptide substrate (Pro-Pro-Gly)₁₀ was hydroxylated less efficiently by the recombinant *D. melanogaster* C-P4H than by the human enzymes. Interestingly, the human α (I) subunit forms an active enzyme tetramer with the *D. melanogaster* PDI, whereas the *D. melanogaster* PH4 α MP assembles only very inefficiently with the human PDI (Annunen *et al.* 1999). As in the case of the assembly of the *C. elegans* mixed tetramer, these results indicate that certain specific structural differences between the α and PDI/ β subunits from various species determine their assembly properties.

We have also expressed a second *D. melanogaster* α subunit, PH4 α EFB, as a recombinant protein in insect cells (unpublished). This forms an active tetramer with *D. melanogaster* PDI and efficiently hydroxylates (Pro-Pro-Gly)₁₀. Unlike PH4 α MP, PH4 α EFB has a histidine residue in the position corresponding to His501 in the human α (I) subunit, and the K_m value of the (PH4 α EFB)₂/(PDI)₂ tetramer for 2-oxoglutarate is similar to those of the two human enzymes (unpublished).

It has been shown very recently that the two *D. melanogaster* salivary gland-specific C-P4H α subunit-related polypeptides, PH4 α SG1 and PH4 α SG2, are necessary for normal salivary gland secretion and the maintenance of an open salivary gland lumen (Abrams *et al.* 2006). As the embryonic salivary glands do not express collagens, it was suggested that these polypeptides are involved in the hydroxylation of some other, as yet unknown secreted or transmembrane proteins (Abrams *et al.* 2006).

2.2.3 Plant and viral prolyl 4-hydroxylases

2.2.3.1 Higher plant and algal prolyl 4-hydroxylases

Plant P4Hs have so far been purified partially or to near homogeneity from the higher plants *Phaseolus vulgaris* and *Vinca rosea* and from the unicellular and multicellular green algae *C. reinhardtii* and *Volvox carterii* (Tanaka *et al.* 1980, Bolwell *et al.* 1985, Kaska *et al.* 1987, 1988). Their analysis has indicated that plant P4Hs are likely to exist as monomers and do not contain PDI as a subunit. They require the same cosubstrates as the animal C-P4Hs, but differ from the latter in that they act primarily on poly(L-proline)-like sequences and hydroxylate collagen-like sequences very inefficiently, if at all (for a review, see Kivirikko *et al.* 1992). Three plant P4Hs have been cloned up to now, two from *A. thaliana* and one from *Nicotiana tabacum*, and these have been shown to be active monomers when characterized as recombinant proteins (Hieta & Myllyharju 2002, Tiainen *et al.* 2005, Yuasa *et al.* 2005).

The genome of *A. thaliana* encodes at least six polypeptides that show 21-27% identity to the catalytic C-terminal regions of the human C-P4H α (I) and α (II) subunits (Hieta & Myllyharju 2002). Two of these, referred to as At-P4H-1 and At-P4H-2, have been cloned and characterized (Hieta & Myllyharju 2002, Tiainen *et al.* 2005). The catalytically critical residues are conserved in them, although the residue in At-P4H-1 that corresponds to the His501 of the human α (I) subunit is an arginine. The At-P4Hs show no sequence similarity to the peptide-substrate-binding domains of the animal C-P4H α subunits and it is still unclear whether plant P4Hs have a separate domain that becomes bound to the peptide substrate or whether peptides interact directly with the catalytic domain. At-P4H-1 is a 283-amino-acid polypeptide, while At-P4H-2 is slightly longer, 299 residues, the amino acid sequence identity between the two being 33%. A distinct difference between these two polypeptides is found in their C-terminal parts, where At-P4H-2 has a 49-amino-acid extension that contains 6 cysteines. A similar C-terminal cysteine-rich extension that is homologous to a 35-residue ShK potassium channel toxin from the sea anemone *Stichodactyla helianthus* (Dauplais *et al.* 1997, Pan *et al.* 1998) is also found in the C terminus of a putative rice P4H (Hieta & Myllyharju 2002, Tiainen *et al.* 2005).

As expected, At-P4H-1 and At-P4H-2 effectively hydroxylated poly(L-proline), the K_m values for poly(L-proline) M, 5000 being 2 and 30 μ M, respectively, and also several plant HRGP-like synthetic peptides (Hieta & Myllyharju 2002, Tiainen *et al.* 2005). Differences in substrate specificity were found between them, however, suggesting that may have different roles *in vivo*. Surprisingly, both enzymes also hydroxylated the collagen-like peptide (Pro-Pro-Gly)₁₀, and At-P4H-1 in particular did this relatively effectively, as reflected in its K_m value of 60 μ M, which is only slightly higher than those of animal C-P4Hs. Furthermore, At-P4H-1 resembled the animal C-P4Hs in that it preferentially hydroxylated the proline preceding glycine in the repeating Pro-Pro-Gly triplets in an asymmetrical manner, the 9th position being most readily hydroxylated. The preference of At-P4H-1 for Y-position hydroxylation was not absolute, however, unlike the situation in the animal C-P4Hs. Both At-P4Hs also hydroxylated a HIF α -like peptide that has a single proline residue, which suggests that, contrary to previous assumptions,

plant P4Hs may not require a type II poly(L-proline) conformation in their substrate for hydroxylation to occur, but rather show sequence specificity (Hieta & Myllyharju 2002, Tiainen *et al.* 2005).

P4Hs have been partially purified from the green algae *C. reinhardtii* and *V. carteri* with purification levels of about 1000 and 7-fold, respectively (Kaska *et al.* 1987, 1988). Both algal enzymes effectively hydroxylated poly(L-proline), but a collagen-like (Pro-Pro-Gly)₁₀ peptide was hydroxylated only to a small extent by the *C. reinhardtii* enzyme, while hydroxylation of this peptide with the *V. carteri* enzyme was not measured. Free proline and the dipeptides Ser-Pro or Ala-Pro were also tested as possible substrate candidates for the *V. carteri* P4H and found not to serve as substrates. The K_m values of both algal enzymes for the reaction cosubstrates resembled those of the vertebrate type I C-P4H with the exception that the K_m for Fe^{2+} was about 10-fold. Overall, the K_m values of the algal enzymes for the reaction cosubstrates and the poly(L-proline) substrate, and also the inhibition constants for certain 2-oxoglutarate analogues, were very similar to each other, the only distinct difference being the 6-fold higher K_m of the *V. carteri* P4H for 2-oxoglutarate. Gel filtration analyses indicated that the apparent molecular weights of the algal enzymes were about 40 kDa (Kaska *et al.* 1987, 1988).

2.2.3.2 *Paramecium bursaria Chlorella virus-1 prolyl 4-hydroxylase*

The genome of PBCV-1 encodes a polypeptide of 242 amino acids that is 15-23% identical to the corresponding C-terminal residues of the human C-P4H $\alpha(I)$ and $\alpha(II)$ subunits, *D. melanogaster* PH4 α MP and *C. elegans* PHY-1 (Eriksson *et al.* 1999). The recombinant PBCV-1 P4H is a soluble monomer that hydroxylates poly(L-proline) and several synthetic peptides that correspond to proline-rich repeats coded by the viral genome. The collagen-like peptide (Pro-Pro-Gly)₁₀ was also hydroxylated by PBCV-1 P4H, although very inefficiently, the K_m value being almost 3 mM. Analysis of the hydroxylation pattern generated in the synthetic peptide (Pro-Ala-Pro-Lys)₅ showed that prolines in both positions were hydroxylated, but those preceding the alanine were hydroxylated more readily. The hydroxylation pattern was symmetric and thus differed from that generated by animal C-P4Hs in (Pro-Pro-Gly)_n peptides (Eriksson *et al.* 1999).

2.3 *Caenorhabditis elegans* and *Chlamydomonas reinhardtii* as model organisms

2.3.1 *C. elegans*

Sydney Brenner stated in 1963 that “the new major problem in molecular biology is the genetics and biochemistry of control mechanisms in cellular development”, and to address the problem he proposed that a “simplest differentiated organism” should be studied “which has a short life cycle, can be easily cultivated and is small enough to be handled in large numbers.” As a result of all those needs, the nematode *C. elegans* rose

quickly to the centre stage in biological research. (for reviews, see Culetto & Sattelle 2000, Jones *et al.* 2005)

There were several reasons that favoured the choice of *C. elegans* as a model organism. One attractive feature is the small size of the worms. The adults are only about 1 mm in length with a maximum diameter of about 80 μm , which means that a large number of worms can be maintained at the same time in a small space. They are also easily cultivated in liquid media or on Petri dishes, with *E. coli* as the only food source. It has a short hermaphroditic lifecycle of only 3.5 days at 20°C. A single adult hermaphrodite can give birth to a large number of progeny, over 300, by self-fertilization. The transparent body of *C. elegans* allows its internal structures and biological processes to be viewed by conventional microscopy, and processes can be observed and explained at the cell level because there are only 959 somatic cells in the adult hermaphrodite. Also, the number of cells is invariant, which has enabled complete description of the cell lineage (for reviews, see Culetto & Sattelle 2000, Jones *et al.* 2005, Kaletta & Hengartner 2006). Drs. Sydney Brenner, H. Robert Horvitz and John E. Sulston were awarded the Nobel Prize in 2002 for their fundamental research into “genetic regulation of organ development and programmed cell death” using *C. elegans* as a model.

The genome of *C. elegans*, about 97 Mb, was the first completely sequenced genome from a multicellular organism (*C. elegans* Genome Consortium 1998). The complete genome sequence of *C. briggsae*, a close relative of *C. elegans*, is now also available, and comparison of the two has indicated that only 4% of the *C. elegans* genes are unique relative to *C. briggsae* (Gupta & Sternberg 2003, Stein *et al.* 2003). On the other hand, about half of the *C. elegans* genes do not share similarity with the genes of any organisms outside the Nematoda phylum. Such genes can be potential drug targets for the prevention and treatment of diseases caused by nematode pathogens both in animals and plants. About 43% of the *C. elegans* genes have a human homologue, including numerous disease-associated genes. As the worms suffer from many diseases that are similar to human ones, such as cancer, neurodegeneration, infectious diseases, diabetes and obesity, and disorders of physiological control and ageing, they can be used to study the basic molecular genetic mechanisms underlying many human disorders and to gain insights into human diseases and into normal function (Barr 2003, O’Kane 2003, Kaletta & Hengartner 2006).

As described above, *C. elegans* has numerous favourable aspects as an experimental model organism but there are also a few weaknesses. *C. elegans* does not have as many specialized tissues as the vertebrates do, and its genes are often highly divergent from their mammalian homologues at the sequence level. The level of divergence reflects the faster rate of sequence change in the worms and does not point to any particular time point at which the separation of the species occurred during evolution. Also, the basic anatomy of the worm is much simpler: its circulatory system, skeletal organization, immune response and cognitive ability are organized in a manner that clearly differs from those of complex organisms. Despite these differences, most of the basic biological processes have remained unchanged throughout evolution and are conserved; e.g. the gene expression machinery, membrane traffic, cytoskeleton, extracellular matrix, cellular asymmetry, epithelial organization, neuronal connectivity, synapse function, cell-cell and intracellular pathways and apoptosis. (for reviews, see Barr 2003, O’Kane 2003)

To understand the biological role of a functional gene, one needs to know the site and timing of its activation and the phenotypic consequences of altering or removing it. Certain techniques developed to study gene expression and function in *C. elegans* have revolutionized biology lately, i.e. RNAi, used to knock down the function of a gene of interest, and green fluorescent protein (GFP) as an expression marker. Andrew Z. Fire and Craig C. Mello developed double-stranded RNA (dsRNA)-dependent posttranscriptional gene silencing methodologies using *C. elegans* (Fire *et al.* 1998) and were granted the Nobel Prize in 2006. Within a few years this technique had been extended to the fly and also to mammalian cells. It is systemic in worms and plants. It is also heritable in *C. elegans*, and the effect can be maintained for over 20 generations and be transmitted with a single sperm (for a review, see Grishok 2005, Vastenhouw *et al.* 2006). dsRNA can be experimentally introduced into *C. elegans* in several ways: by direct injection into the gonad or body cavity, by soaking the nematodes in dsRNA, by ingestion of bacteria that are engineered to produce dsRNA or by generating heritable inverted repeat genes. The introduction of dsRNA triggers degradation of the endogenous mRNA through a specific nuclease pathway and as a consequence the gene of interest becomes inactivated in virtually all cells. Induction of RNAi in *C. elegans* requires a dsRNA with a minimal length of 100 bp. This capacity to use long stretches of dsRNA which gives rise to many different small interfering RNA (siRNA) molecules, increases the efficiency of RNAi. On the other hand, the multitude of siRNAs may also increase the risk of inactivating other genes. The introduction of dsRNA into the body cavity of the worm and its spread there will develop a phenotype as soon as the activity of the gene product is significantly reduced. A major advantage of RNAi lies in the flexibility of timing: it can be induced at any time during the life cycle. This offers an opportunity to study the function of a gene at any stage of interest (Fire *et al.* 1998, Timmons *et al.* 2001, Timmons *et al.* 2003, Hariharan & Haber 2003, Grishok 2005, Kaletta & Hengartner 2006). Numerous genome-wide RNAi projects concerned with *C. elegans* are currently going on (Fraser *et al.* 2000, Gönczy *et al.* 2000, Maeda *et al.* 2001, Kamath *et al.* 2003, Simmer *et al.* 2003, Frand *et al.* 2005, Clark & Ding 2006, Lappe & Roy 2006).

Two other significant techniques in the field of *C. elegans* research are the use of GFP as a marker and micro-array technology in gene expression studies. GFP has become a standard fluorescent reporter for gene expression or subcellular protein localization, although the technique relies largely on the cell identification ability of the microscopist. Micro-array analysis of gene expression levels has provided new approaches for tissue-specific gene profiling, for examining the genome organization and for exploring overall changes in gene expression (for reviews, see Barr 2003, Kaletta & Hengartner 2006).

Large-scale projects for sequencing the genomes of different organisms have greatly facilitated the identification of specific genes and the elucidation of their roles *in vivo*. The sequence data are saved in databases that are freely available for everyone through the Internet, thus enabling researchers worldwide to search a vast content of highly curated sequence data. The Wormbase (<http://www.wormbase.org/>) for *C. elegans* is one of the web-based model organism databases that is maintained by the International Wormbase Consortium. It originated in 2001 from its predecessor AceDB and has evolved and expanded since then through the development and utilization of better software that provides renewable data on classical genetics, cell biology and functional

genetics, and also genome sequences for other closely related nematode species. (Bieri *et al.* 2006)

2.3.2 *C. reinhardtii*

The first studies performed on the green alga *Chlamydomonas* go back to the early twentieth century when the first morphological descriptions and genetic studies were made. A. Pascher demonstrated the suitability of various *Chlamydomonas* species for genetic analysis in 1918 and pointed out the advantages of a haploid system, while Franz Moewus demonstrated in the 1930s that *Chlamydomonas* mutants can be isolated and characterized, and the use of these species as laboratory organisms was developed later, in the 1940s and 1950s. The principal laboratory strains of *C. reinhardtii* are thought to be derived from isolates obtained from soil in 1945. (for reviews, see Lefebvre & Silflow 1999, Harris 2001)

Chlamydomonas is a unicellular organism with a simple life cycle. The wild-type *C. reinhardtii* cell is about 10 μm in diameter and is enclosed within a wall consisting primarily of HRGPs. It contains all the organelles common to plant cells, including several mitochondria and a single chloroplast, and additional organelles such as the two flagella that are crucial for its motility. *Chlamydomonas* cells can be grown easily in defined solid and liquid medium at neutral pH and have no extra requirements for supplementary vitamins or other cofactors. Depending on the strain and culture conditions, they grow fast, with a doubling time of about 5-8h (for a review, see Nickelsen & Kuck 2000). In addition, they have a capacity to grow with light as the sole energy source, or on acetate in the dark. The cells grown in the dark maintain a normal chloroplast structure and pigmentation and resume photosynthetic CO_2 fixation upon illumination. This has enabled studies of photosynthesis and the isolation of a broad range of gene mutations that affect the photosynthetic function. Many of these *Chlamydomonas* genes are probably important for chloroplast assembly and function in higher plants as well (for reviews, see Lefebvre & Silflow 1999, Harris 2001, Grossman *et al.* 2004). *C. reinhardtii* cells are normally haploid and the species has two genetically fixed mating types. Because of this haploidism, mutant phenotypes can be identified directly after mutagenesis without any need for further genetic manipulations (for a review, see Nickelsen & Kuck 2000).

C. reinhardtii contains three autonomous genetic systems: the nucleus, chloroplast and mitochondria. Both the chloroplast and nuclear genomes are readily transformed by a variety of methods (for a review, see Grossman *et al.* 2004). Transformation of the nuclear genome is typically carried out using particle bombardment, glass beads, carbide whisker agitation or electroporation (Franklin & Mayfield 2004). The integration of transforming DNA into the nuclear genome is a random process in *C. reinhardtii* and higher plants and results in a variety of expression levels that are caused by position effects on the integrating DNA (for a review, see Grossman *et al.* 2003). Eukaryotic algae have been considered a novel option as an inexpensive method for the large-scale production of recombinant proteins, as they can be cultivated in a cheap and easy manner and grown to high cell densities (Griesbeck *et al.* 2006). Although a complete toolkit for

the genetic manipulation of *C. reinhardtii* is now available, the expression of heterologous genes in it is still somewhat problematic. The nuclear genome, with a size of 100 to 110 Mb, has a GC content of 64% and almost 85% of each third nucleotide is a G or C. By optimizing the heterologous gene to match this codon usage, it is possible to increase the expression levels of foreign proteins. Another factor that can increase transgene expression is the inclusion of introns with the sequence to be expressed. The first intron from the *C. reinhardtii* gene, which encodes the small subunit of ribulose 1,5-bisphosphate carboxylase (*rbcS2*), acts as an enhancer-like element and has been shown to increase the level of phleomycin expression, for example, in a manner that is dependent on its orientation or position relative to the *RBCS2* promoter used to drive expression of the heterologous gene. Thus high levels of GFP expression have been achieved when codon-optimized GFP with multiple introns is driven by an *RBCS2* promoter. These high levels are quickly suppressed in the transgenic strain, however, and it seems that *C. reinhardtii* has an ability to invoke gene-silencing mechanisms upon transgene expression (Franklin & Mayfield 2004).

The generation of a comprehensive *Chlamydomonas* EST library covering about 200,000 ESTs and the sequencing of its nuclear genome by the US Department of Energy Joint Genome Institute has enabled reverse genetics approaches to be adopted. RNA silencing using constructs that express antisense or inverted repeat-containing RNA has led to the successful knock-down of a gene of interest in *Chlamydomonas*. One question that is currently under investigation is whether RNAi-mediated knock-down in *Chlamydomonas* requires the same components as in *C. elegans*, for example, and sequences have been identified in the algal genome that are homologous to those known to encode components needed in the *C. elegans* RNA silencing pathway. (Schroda 2006)

3 Outlines of the present research

All plant P4Hs characterized so far have been shown to be monomers, and they would thus be interesting targets for structural studies instead of multimeric animal C-P4Hs, which have turned out to be difficult to crystallize. Unfortunately, the two previously cloned plant P4Hs, from *A. thaliana*, are poorly soluble when expressed as recombinant proteins in insect cells and are therefore not suitable for structural studies. As a soluble monomeric P4H has previously been partially purified from the green alga *C. reinhardtii*, we set out to:

1. clone a P4H from *C. reinhardtii*, express and purify it as a recombinant protein, determine its catalytic properties, and study its function *in vivo* using RNAi.

C-P4Hs have been shown to form enzyme families in all animal species studied, from vertebrates to invertebrates. Three C-P4H α subunit isoforms from *C. elegans* have been characterized so far. The aim of the second part of this work was to complete the characterization of the *C. elegans* C-P4H family and to study the cuticle C-P4Hs of a close relative, *C. briggsae*. The specific aims were:

2. to characterize the remaining two *C. elegans* C-P4H α subunit-like genes, study their expression patterns and *in vivo* functions, and clone the corresponding cDNAs and express the α subunits as recombinant proteins in order to study their catalytic properties, and
3. to clone and characterize *C. briggsae* PHY-1 and PHY-2 homologues, compare their assembly properties with those of the corresponding *C. elegans* polypeptides, and analyze the inter-species compensatory potential of their genes.

4 Materials and methods

The materials and methods used in this thesis are summarized in Tables 1 and 2 below. Detailed descriptions of the experimental procedures, including the references, are available in the original articles I-III.

Table 1. Strains.

Strain	Origin
<i>C. elegans</i> Bristol N2	Caenorhabditis Genetics Center
<i>C. elegans daf-2(e1370)</i>	Caenorhabditis Genetics Center
<i>C. elegans</i> NL2099	Caenorhabditis Genetics Center
<i>C. elegans</i> CB364, <i>dpy18 (e364)</i>	Caenorhabditis Genetics Center
<i>C. elegans dpy18 (e1096)</i>	Caenorhabditis Genetics Center
<i>C. elegans phy-2</i>	Caenorhabditis Genetics Center
<i>C. elegans phy-3</i>	Caenorhabditis Genetics Center
<i>C. elegans</i> AF16	Caenorhabditis Genetics Center
<i>C. elegans dpy18</i>	Generated in this studywork
<i>C. elegans</i> CC125 mt+ 137c	Chlamydomonas Genetics Center
<i>C. elegans</i> CC620 mt+ 137c	Chlamydomonas Genetics Center
<i>C. elegans</i> CC621 mt- 137c	Chlamydomonas Genetics Center

Table 2. Methods.

Level	Method	Original publication
DNA	Isolation of genomic DNA	I
	Rapid amplification of cDNA ends (RACE)	II
	Cloning techniques	I, II, III
	PCR	I, II, III
	Site-directed mutagenesis	III
	Generation of a mutant <i>C. briggsae</i> strain	III
	LacZ-reporter gene expression analysis	II
RNA & protein	Expression and analysis of recombinant proteins in insect cells	I, II, III
	Expression and analysis of recombinant proteins in <i>E. coli</i> .	I
	SDS-PAGE, native PAGE and Western blotting	I, II, III
	Purification of recombinant proteins	I
	Gel filtration	I, II, III
	N-terminal sequencing	I
	Immunofluorescence staining	II
	RNA isolation and RT-PCR	I, II, III
	Micro-array analysis	II
	RNA interference	I, II, III
	Circular dichroism spectroscopy	I
Others	P4H activity assays	I, II, III
	Preparation of autolysin	I
	Microinjection	II, III
	Electroporation	I
	Transmission electron microscopy	I

5 Results

5.1 Characterization of the *C. reinhardtii* prolyl 4-hydroxylase gene family and detailed analysis of one of the encoded polypeptides, Cr-P4H-1

A sequence homology search of the *C. reinhardtii* EST database indicated the presence of two overlapping EST sequences (accession numbers AW661167 and BG860906) that encode a 253-residue polypeptide with similarity to the C-terminal catalytic part of the human C-P4H α (I) subunit (Figure 1A in I). The corresponding full-length cDNA was cloned, its 5' end was verified by 5'RACE reactions, and the encoded polypeptide was named Cr-P4H-1. In addition, a sequence homology search of the *C. reinhardtii* genome draft release, version 2.0, identified the presence of 10 putative genes coding for P4H-like polypeptides (Figure 1B in I). One of them, a polypeptide encoded by a gene with the annotation C_110025, was identical to Cr-P4H-1 with the exception that the three C-terminal residues of Cr-P4H-1 were replaced by a cysteine-rich stretch of 47 residues (Figure 1B in I). This indicates that the gene for Cr-P4H-1 is alternatively spliced, so that the short and long forms of the polypeptide were named Cr-P4H-1A and Cr-P4H1B, respectively. The C-terminal cysteine-rich stretch of Cr-P4H-1B is homologous to the 35-residue sea anemone ShK potassium channel toxin (Dauplais *et al.* 1997, Pan *et al.* 1998) and is also found in the previously cloned At-P4H-2 (Tiainen *et al.* 2005). The other nine polypeptides, encoded by genes with the annotations C_360012, C_390100, C_150026, C_280045, C_1000010, C_2500005, C_50128, C_130209 and C_800043, varied in length from 253 to 376 residues, and eight of them had a similar C-terminal cysteine-rich extension to Cr-P4H-1B (Figure 1B in I). The 253-residue polypeptide coded C_360012 had the highest amino acid sequence identity to Cr-P4H-1A, 41%. Eight of the polypeptides, including both forms of Cr-P4H-1, were predicted to have a cleavable signal peptide, while three of them had a non-cleavable one. Catalytically important amino acids present in the C-terminal part of the polypeptides were conserved in all the sequences, with only a few exceptions. The second iron binding histidine in C_800043 was replaced by a glutamate and the fifth critical histidine in C_2500005 by an arginine (Figure 1B in I). The latter is likely to be a functional P4H, as an arginine in this position

is found in *D. melanogaster* PH4 α MP, for example (Annunen *et al.* 1999), and in At-P4H-1 (Hieta & Myllyharju 2002). Whether the polypeptide encoded by the gene C_800043 is an active P4H is more unpredictable, as substitution of the second iron binding histidine in bovine aspartyl (asparaginy) β -hydroxylase results in 10-20% residual activity (McGinnis *et al.* 1996), while the corresponding mutation in the human C-P4H α (I) subunit generates an inactive enzyme tetramer (Myllyharju & Kivirikko 1997).

5.1.1 Amino acid sequence of C. reinhardtii P4H-1

The 253-amino-acid Cr-P4H-1A has a cleavable signal peptide of 16 residues and is 26% identical to the catalytic C-terminal part of the human α (I) subunit, but it shows a higher degree of identity, 40%, with At-P4H-1 and At-P4H-2 and with the PBCV-1 P4H, 30% (Figure 1A in I). All five catalytically critical residues are conserved in Cr-P4H-1A and B (Figure 1A in I). Cr-P4H-1A contains three cysteine residues in conserved positions as compared with At-P4H-1 (Figure 1A in I). Phylogenetic relationships between the amino acid sequences of plant P4Hs and animal C-P4H α subunits from different species are shown in Figure 2 of the original article I. Phylogenetic analysis showed that Cr-P4H-1 is evolutionarily closest to At-P4H-1 and forms a clade with At-P4H-1 and At-P4H-2, and these three P4Hs are in general closely related to other short P4Hs and C-P4H α subunits, i.e. PBCV-1 P4H and *C. elegans* PHY-3, respectively (Figure 2 in I). Of the longer C-P4H α subunits, the Cr-P4H-1 and At-P4Hs are more closely related to the vertebrate α (III) subunits than to the α (I) and α (II) subunits or *C. elegans* PHY-1 and PHY-2 (Figure 2 in I).

5.1.2 Expression of recombinant Cr-P4H-1A in insect cells

A baculovirus vector carrying the Cr-P4H-1A sequence was cotransfected into *Spodoptera frugiperda* Sf9 cells with BaculoGold DNA, and insect cells were infected with the amplified virus and harvested 72 h after infection. The cells were homogenized in a buffer containing Triton X-100 and centrifuged, the remaining insoluble cell pellets were further solubilized in 1% SDS and aliquots of both samples were analyzed by SDS-PAGE. The majority of the recombinant Cr-P4H-1A polypeptide was found in the insoluble fraction (Figure 3A in I), but a clear band was also seen in the soluble fraction (Figure 3A in I). The recombinant Cr-P4H-1A was purified from the Triton X-100-soluble fraction by gel filtration and analyzed by SDS-PAGE. The results showed that the enzyme was purified to near homogeneity by a single-step gel filtration (Figure 3A in I). Experiments using a calibrated gel filtration column indicated that the enzyme was eluted in fractions corresponding to a molecular weight of \sim 35 kDa. The recombinant Cr-P4H-1A was thus a monomer, since its calculated molecular weight without the signal peptide is \sim 28 kDa.

5.1.3 Expression of recombinant Cr-P4H-1A and Cr-P4H-1B in *E. coli*

E. coli expression vectors coding for Cr-P4H-1A and Cr-P4H-1B with an N-terminal histidine tag were transformed into the *E. coli* Origami™ (DE3) strain and expression was induced with isopropyl-1-thio- β -D-galactopyranoside. The cells were harvested after overnight induction, suspended in a P4H buffer, disrupted by sonication and centrifuged, after which the soluble and insoluble fractions were analyzed by SDS-PAGE. The majority of the Cr-P4H-1A and about half of the Cr-P4H-1B were found in the soluble fractions and the recombinant proteins could be purified essentially to homogeneity by metal affinity chromatography and gel filtration (Figure 3B and C in I). Circular dichroism spectroscopy was used to study the thermal stability of the recombinant Cr-P4H-1A. A cooperative transition of a folded state to an unfolded one was observed, the melting temperature being $\sim 58^\circ\text{C}$, which indicates that the polypeptide has a fairly compact, stable structure (Figure 4 in I).

5.1.4 Catalytic properties of the recombinant Cr-P4H-1A and Cr-P4H-1B

The catalytic activity of purified recombinant Cr-P4H-1A produced in insect cells and *E. coli* and of Cr-P4H-1B produced in *E. coli* was assayed by a method based on the hydroxylation-coupled decarboxylation of 2-oxo[1- ^{14}C]glutarate. Both Cr-P4H-1A and Cr-P4H-1B showed activity when poly(L-proline) was used as a substrate and the specific activities of both variants were about 300 mol/mol enzyme/min. As no differences were found in the specific activities of the two variants, Cr-P4H-1A was used in all the subsequent kinetic analyses. The K_m value for Fe^{2+} , 30 μM (Table 1 in I), resembled that of recombinant At-P4H-1 (16 μM , Hieta & Myllyharju 2002), but was about 6-fold higher than that of At-P4H-2 (Tiainen *et al.* 2005) and 75-fold higher than that of the PBCV-1 P4H (Eriksson *et al.* 1999). The K_m value for 2-oxoglutarate, 250 μM (Table 1 in I), was 1.5-2-fold higher than those of the two At-P4Hs and about 10-fold relative to that of the PBCV-1 P4H, while the K_m for ascorbate, 20 μM (Table 1 in I), was about 15-fold lower than the values reported for the two At-P4Hs and the PBCV-1 P4H.

The inhibitory effects of the 2-oxoglutarate analogues pyridine-2,4-dicarboxylate and pyridine-2,5-dicarboxylate were also tested, since both are known to be effective competitive inhibitors of vertebrate C-P4Hs, and they were also found to inhibit Cr-P4H-1A (Table 1 in I). The K_i values were 70 μM and 100 μM , respectively, thus being 35-fold and 125-fold higher than the corresponding values for human type I C-P4H (Kivirikko *et al.* 1992). These inhibitory properties of Cr-P4H-1A resembled more closely those of the plant enzyme At-P4H-2, the K_i values of which for these two inhibitors are only 7-fold and 1.3-fold lower, respectively (Tiainen *et al.* 2005). Zn^{2+} was found to inhibit Cr-P4H-1A relatively effectively with respect to Fe^{2+} , the K_i value being 10 μM (Table 1 in I), which is markedly higher than that of the human type I C-P4H (Kivirikko *et al.* 1992).

In addition to poly(L-proline), Cr-P4H-1A hydroxylated several synthetic peptides representing proline-rich sequences present in *C. reinhardtii* HRGPs (Table 2 in I).

Domain 3 of the *C. reinhardtii* cell wall protein GP1 has a characteristic proline-rich repeat and a synthetic peptide (Pro-Ser)₁₃-Pro-Ile-(Pro-Ser)₂-Pro-Lys-(Pro-Ser)₂-Pro representing this was the most effectively hydroxylated one among the HRGP peptides tested (Table 2 in I). The K_m value for this peptide was 120 μ M and the V_{max} obtained with it was identical to that obtained with poly(L-proline), M_r 5,000, but 10% lower than the highest V_{max} obtained with poly(L-proline), M_r 30,000 (Table 2 in I). Three other proline-rich peptides, (Ser-Pro-Ala/Glu/Lys-Pro-Pro)₅, representing the motif -Pro-Pro-Ser-Pro-X- found in domain 2 of GP1, the sexual agglutinins Sag1 and Sad1, and a protein A2 of unknown function, also served as substrates for Cr-P4H-1A, the K_m value being identical with all three, 100 μ M (Table 2 in I), although the V_{max} values varied from 25% to 65% of that obtained with poly(L-proline), M_r 30,000 (Table 2 in I). The 75% reduction in V_{max} obtained with the peptide (Ser-Pro-Lys-Pro-Pro)₅ may reflect the observations that a proline following a lysine is never hydroxylated in higher plants whereas a proline preceding a lysine is hydroxylated (Kieliszewski & Lamport 1994).

Cr-P4H-1A also hydroxylated the collagen-like (Pro-Pro-Gly)₁₀ peptide, but only inefficiently, with a K_m higher than 1.5 mM and a V_{max} 60% of that obtained with poly(L-proline) (Table 2 in I). It thus resembles more closely At-P4H-2 and the PBCV-1 P4H than At-P4H-1 in this respect. In order to determine which proline residues of the (Pro-Pro-Gly)₁₀ peptide are preferentially hydroxylated by Cr-P4H-1A, a partially hydroxylated peptide was purified from the reaction mixture and analyzed by N-terminal sequencing. Interestingly, the highest degree of hydroxylation was found in the proline residues present in positions following a glycine residue, while the prolines preceding the glycines were hydroxylated to a lesser degree (Figure 5 in I). The hydroxylation pattern thus differs greatly from those of At-P4H-1 and the vertebrate C-P4Hs, the former acting preferentially and the latter exclusively on the prolines preceding a glycine.

5.1.5 Analysis of the in vivo function of Cr-P4H-1 by RNAi

The *in vivo* role of Cr-P4H-1 was studied by RNAi, a mechanism discovered by Andrew Fire and Craig Mello in *C. elegans* in 1998 in which the function of a gene of interest is suppressed by introducing double-stranded RNA into a cell. dsRNA is cleaved by the enzyme Dicer to small 21-bp RNA (siRNA) stretches that bind with the help of the multi-enzyme RNA-induced silencing complex to the target mRNA, which is eventually degraded (Fire *et al.* 1998). Regarding the possibility of unspecific silencing the existence of homologous 21-bp sequences among the 10 P4H-like genes present in the *C. reinhardtii* genome was analyzed. One of the sequences, C_360012, contained one 21-bp region with three mismatches relative to Cr-P4H-1 located in the centre of the sequence while all the other 21-bp regions of the Cr-P4H-like sequences had more than three mismatches. Target recognition between siRNA and mRNA is highly specific and a mismatch in even a single nucleotide has been shown to cause a substantial decrease in RNAi efficiency (Elbashir *et al.* 2001). Our analysis thus indicated that only expression of Cr-P4H-1 can be expected to be affected in RNAi experiments.

The hairpin dsRNA technique was used to suppress expression of the Cr-P4H-1 gene. The *RBCS2* promoter and the dsRNA construct of the Cr-P4H-1 gene were fused to a

pSP124S plasmid that contains the phleomycin resistance gene, *ble*, as a dominant marker for nuclear transformation (Figure 6A in I). The RNAi plasmid or the empty pSP124S plasmid were transformed by electroporation into autolysin-treated *C. reinhardtii* wild-type CC125mt+ 137C cells. Independent RNAi and control lines were generated from single transformant clones selected in Zeocin plates and the existence of the RNAi construct in the RNAi transformants was verified by PCR analysis (Figure 6B in I).

The *C. reinhardtii* cell wall is seen by EM to be composed of five structurally distinct layers. The innermost and outermost layers, W1 and W7, are composed of a loose network of fibres, the dense W2 and W6 layers having a granular W4 layer between them (Roberts *et al.* 1972, Goodenough & Heuser 1985, Adair & Snell 1990). The cell wall structures of wild-type, control and two independent RNAi transformant clones were analyzed by transmission electron microscopy of high-pressure frozen and freeze-substituted samples. The wild-type and control cells had a well-defined cell wall with distinct layers and the plasma membrane was closely affixed to this wall (Figure 7A and C in I), whereas in the majority of the RNAi transformants the cell wall structure was abnormal (Table 4 in I), in that the cell wall was less clearly defined and was also detached from the plasma membrane in some cases (Figure 7B and D-F in I). These cells lacked a multilayered structure and seemed instead to be entirely composed of a network of fibrils resembling those of the W1 and W7 layers (Figure 7 D-F in I). Quantitative analysis showed that 95% of the control line cells had a normal cell wall structure with a closely affixed plasma membrane, while in two RNAi transformant lines 84-88% of the cells had a clearly altered cell wall and/or detachment of the plasma membrane (Table 4 in I). The effect of Cr-P4H-1 RNAi was also studied at the protein level. Cell extracts from three independent control and four RNAi lines were analyzed for P4H activity and the activity levels in the RNAi were found to have decreased to 33-58% of that of the controls (Table 3 in I).

The growth of several RNAi transformant and control lines was also analyzed. Growth rates were followed for 10 days and the cell densities were calculated using a haemocytometer. No clear differences between the growth rates of the control and Cr-P4H-1 RNAi transformant lines were observed, indicating that the lack of Cr-P4H-1 activity has no effect on growth. Previous analyses have shown that the growth properties of various isolates of a cell wall defective mutant of a related species *C. monoica* are variable: some of them show a faster doubling time than wild-type cells, while others grow more slowly (Fuentes & Van Winkle-Swift 2003), indicating that cell wall defects do not always lead to growth problems.

5.2 Cloning and characterization of a novel *C. elegans* prolyl 4-hydroxylase

5.2.1 Molecular cloning of the *C. elegans phy-4*

A sequence homology search of the *C. elegans* genome indicated the presence of two predicted open reading frames from genes Y43F8B.4 and C14E2.4 that code for polypeptides having homology with the C-terminal regions of the human C-P4H α (I) and α (II) subunits and *C. elegans* PHY-1, PHY-2 and PHY-3. The Y43F8B.4 gene was predicted to contain seven exons coding for a 533-amino-acid polypeptide with two sets of catalytically critical residues (Figure 1A in II) and an N-terminal transmembrane region with a low probability of cleavage, indicating that it has a non-cleavable signal peptide. Cloning of the cDNA representing this gene using PCR with exon-specific primers and 5' and 3' RACE reactions showed that three transcripts are generated from the Y43F8B.4 gene (Figure 1A in II). Two of the transcripts, termed *phy-4.1* and *phy-4.2*, contain the exon 1, 2 and 3 sequences as predicted, but differ in their 3' ends (Figure 1A in II). In *phy-4.1* the sequence continues from exon 4 to intron 4, where an in-frame stop codon is located at nucleotides 65-67 of the intron 4 sequence (Figure 1A in II). In *phy-4.2* exon 4 is spliced to exon 5, but so that the last nucleotide of exon 4 is deleted, causing a frame-shift and a premature stop codon in exon 5 (Figure 1A in II). The third transcript, *phy-4.3*, covers the 3' end of exon 5 and exons 6 and 7, which are spliced together so that four nucleotides are lost, causing a frame-shift and a premature stop codon (Figure 1A in II). Due to the deletion, the transcript no longer codes for the catalytically important amino acids and the encoded polypeptide cannot be a functional α subunit (Figure 1A in II). The existence of the predicted full-length Y43F8B.4 transcript was excluded in our analysis.

The PHY-4.1 and PHY-4.2 polypeptides, starting from Leu18 after the transmembrane region, are 23-25% and 21-26% identical to the corresponding residues of PHY-1, PHY-2 and PHY-3 (Figure 2 in II). The highest amino acid sequence identity is found around the catalytically critical residues, the amino acids 138-256 of PHY-4.1 and PHY-4.2 being 35-37% identical to the corresponding regions of the other PHY polypeptides (Figure 2 in II). PHY-4.1 has four cysteines in identical positions to those of PHY-1 and PHY-2 (Figure 2 in II). The corresponding conserved cysteines have been shown to be essential for intrachain disulphide bond formation in the human C-P4H α (I) subunit (John & Bulleid 1994, Lamberg *et al.* 1995). The extreme C-terminal one of these conserved cysteines is not present in PHY-4.2 (Figure 2 in II). Both PHY-4.1 and PHY-4.2 contain one potential N-glycosylation site (Figure 2 in II).

C14E2.4 was predicted to contain six exons coding for a 429-amino acid polypeptide. Sequence analysis of the cDNA products showed that a 4-nucleotide deletion in the end of exon 4 leads to a frame-shift and a lack of codons for two catalytically critical residues (Figure 1B in II). Consequently, C14E2.4 does not code for a functional PHY isoform.

5.2.2 Expression of recombinant PHY-4.1 and PHY-4.2 in insect cells

Recombinant baculoviruses coding for the PHY-4.1 and PHY-4.2 polypeptides with a baculoviral GP67 signal peptide were generated and used to produce recombinant proteins in insect cells. The cells were harvested 72 h after infection, homogenized in a buffer containing 0.1% Triton X-100 and centrifuged. The pellets were further solubilized in 1% SDS, and the Triton X-100 and SDS-soluble samples were analyzed by 8% SDS-PAGE and non-denaturing PAGE followed by Coomassie Blue staining or Western blotting. The major parts of both PHY-4 forms were not soluble in the Triton X-100 buffer and could be solubilized only with 1% SDS (Figure 3A in II). The results of the N-glycosidase F treatment and N-terminal sequencing of the recombinant PHY-4.1 showed that it was glycosylated (Figure 3 in II) and the GP67 signal was correctly cleaved, indicating that the polypeptide was correctly translocated into the lumen of the ER.

To study the association of PHY-4.1 and PHY-4.2 with various *C. elegans* PDI isoforms, insect cells were coinfecting with recombinant viruses coding for PHY-4.1 or PHY-4.2 with *C. elegans* PDI-1, PDI-2 and PDI-3. Triton X-100-soluble extracts of the cell homogenates were analyzed by non-denaturing PAGE followed by Western blotting with PHY-4 and PDI-2 antibodies. PHY-4.1 was found to assemble with PDI-2 but not with the other two PDI isoforms, while PHY-4.2 did not associate with any of the PDI polypeptides (Figure 3C and D in II). As seen in the non-denaturing PAGE analysis, PHY-4.1 and PDI-2 formed dimers and tetramers, and this was confirmed by analysis in a calibrated gel filtration column, the formation of dimers being more efficient than that of tetramers (Figure 3C and D in II).

5.2.3 Catalytic properties of the recombinant P4H tetramers and dimers containing PHY-4.1 as the catalytic subunit

Triton X-100 extracts from insect cells coexpressing PHY-4.1 and PDI-2 or PHY-4.2 and PDI-2 were analyzed for C-P4H activity using an assay based on the hydroxylation-coupled decarboxylation of 2-oxo-[1-¹⁴C]glutarate and collagen-like peptide (Pro-Pro-Gly)₁₀ as a substrate. C-P4H activity was generated in the extracts from cells coexpressing PHY-4.1 and PDI-2, but only at relatively high substrate concentrations, while no activity was detected in the extracts from cells coexpressing PHY-4.2 and PDI-2. As the (Pro-Pro-Gly)₁₀ peptide seemed to be a relatively poor substrate for the cell extract containing a mixture of recombinant (PHY-4.1)₂/(PDI-2)₂ tetramers and PHY-4.1/PDI-2 dimers, the hydroxylation of other proline-rich sequences was analyzed. The peptide (Pro-Glu-Pro-Pro-Ala)₅ and poly(L-proline) were also found to act as substrates (Table 1 in II), while many other proline-rich peptides tested were not hydroxylated (see original article II for the peptide sequences).

The K_m value of the mixture of (PHY-4.1)₂/(PDI-2)₂ tetramers and PHY-4.1/PDI-2 dimers for (Pro-Pro-Gly)₁₀ was >2.5 mM, being more than 100-fold higher than that of the main *C. elegans* C-P4H PHY-1/PHY-2/(PDI-2)₂ and about the same as that of the viral PBCV-1 P4H (Table 2 in II). The (Pro-Glu-Pro-Pro-Ala)₅ peptide representing a

proline-rich repeat coded by the PBCV-1 genome was hydroxylated efficiently by the mixture of tetramers and dimers containing PHY-4.1, the K_m value being 4-fold lower than that of the PBCV-1 P4H (Table 2 in II). The K_m values for poly(L-proline), M_r 8900 and 32,000, were also lower than those of the viral enzyme (Table 2 in II).

The K_m values of the (PHY-4.1)₂/(PDI-2)₂ tetramers and PHY-4.1/PDI-2 dimers for reaction cosubstrates were also determined (Table 3 in II). The K_m for Fe²⁺ was five times higher than that of the PHY-1/PHY-2/(PDI-2)₂ tetramer and human type I C-P4H and one-third of that of Cr-P4H-1 (Table 3 in II). The K_m for 2-oxoglutarate was similar to that of the human type I C-P4H, slightly lower than that of the PHY-1/PHY-2/(PDI-2)₂ tetramer, and markedly lower than that of Cr-P4H-1 (Table 3 in II). The K_m for ascorbate was unusually low relative to the other two animal C-P4Hs, being similar to that of Cr-P4H-1 (Table 3 in II).

5.2.4 Analysis of the expression of *phy-4*

Since it had seemed during the cloning of PHY-4.1 that its cDNA was more abundant in cDNA pools generated from heat-shocked nematodes than from those maintained under standard culture conditions, the expression level of *phy-4.1* after heat shock treatment was analyzed in a micro-array hybridization experiment on a chip representing about 22,500 *C. elegans* genes. The expression level of *phy-4.1* was not markedly elevated upon heat shock, however, nor were those of the genes coding for the cuticle C-P4H subunits PHY-1, PHY-2 and PDI-2, or the isoform PHY-3.

Two approaches were used to study the spatial expression of PHY-4.1, reporter gene expression and immunostaining. A putative *phy-4.1* promoter fragment from -953 to +6 relative to the translation initiation codon was ligated in-frame to a *lacZ* reporter gene. The construct was microinjected into the germ line with a marker plasmid that contains the *rol-6(su1006)* gene. A large number of individual nematodes from three independent lines were stained for β -galactosidase activity. Expression of *phy-4::lacZ* was found to be restricted to three pharyngeal gland cell nuclei g1AL, g1AR and g1P (Figure 4A in III) in all larval stages, adults and late embryos. The gland cells fill the space in the terminal bulb of the pharynx and extend into three cuticle-lined ducts anteriorly from the terminal bulb. Two of the three ducts pass through the isthmus before opening into the pharyngeal lumen near the anterior bulb, while one of them extends farther and empties near the anterior limit of the pharynx (Albertson & Thomson 1976).

Expression of PHY-4.1 in the pharyngeal gland cells was confirmed by immunofluorescence staining with a purified polyclonal PHY-4.1 antibody (Figure 4B in III). Additional staining was found at the boundary between the pharynx and the gut and in the excretory duct (Figure 4B in III). The excretory duct is a cuticle-lined tube that lies next to the terminal bulb of the pharynx and is involved in the secretion of metabolites from cells of the excretory system (Nelson *et al.* 1983).

5.2.5 Analysis of the *in vivo* function of *phy-4* by RNAi

The *in vivo* function of *phy-4* was examined using standard RNAi protocols in wild-type, RNAi sensitive NL2099, *dpy-18(e364)*, *phy-2(ok177)* and *phy-3(ok199)* backgrounds. A cDNA fragment corresponding to nt 162-723 of the coding region of the *phy-4.1* and *phy-4.2* cDNAs was amplified from a *C. elegans* mixed-stage cDNA pool, cloned into the vector pPD129:36 and double-stranded RNA for injection was produced *in vitro*. 50 young adults were microinjected, allowed to recover overnight, and the progenies generated were analyzed. To perform bacterially mediated RNAi, the construct pPD129:36-*phy-4* was transformed into *E. coli* and feeding plates were generated from the transformed bacteria. Several L4 animals were transferred to the feeding plates and incubated for 2 days at 15°C, 20°C and 25°C, after which the adults were transferred to fresh feeding plates, allowed to lay eggs for 24 h and the progeny were analyzed.

RNAi of *phy-4.1* did not lead to any obvious defects in the overall morphology, fertility or behaviour of the wild-type or RNAi sensitive strains. No additional phenotypic effects were observed for the *phy-4.1* RNAi in the *dpy-18(e364)* mutant background, the progeny remaining dumpy because of the *phy-1* mutation, nor were any obvious abnormalities generated in the *phy-2(ok177)* or *phy-3(ok199)* mutant nematodes by *phy-4.1* RNAi.

5.3 Analysis of *C. briggsae* cuticle collagen prolyl 4-hydroxylases and comparison with those of *C. elegans*

5.3.1 Cloning of *C. briggsae* PHY-1, PHY-2 and PDI-2

A sequence homology search of the *C. briggsae* genome indicated the presence of genes coding for proteins that are homologous with those of the *C. elegans* cuticle C-P4H subunits PHY-1, PHY-2 and PDI-2, and the corresponding *C. briggsae* cDNAs were cloned. The amino acid sequence identity between the *C. briggsae* and *C. elegans* PHY-1, PHY-2 and PDI-2 is very high, being 92%, 96% and 97%, respectively (Figure 1 in III). All the catalytically important amino acids and cysteines required for intrachain disulphide bonding are conserved in the PHY polypeptides (Figure 1 in III), and the *C. briggsae* PHY-1 also has the 18-amino-acid C-terminal extension found in the *C. elegans* PHY-1 (Figure 1 in III).

5.3.2 Rescue experiments on *C. elegans* and *C. briggsae* *phy-1* mutants using wild-type *phy-1* and *phy-2* genes from both species

A *C. briggsae* *phy-1* mutant was generated using *D. melanogaster* *Mos1* transposon insertion (Besserau *et al.* 2001, Williams *et al.* 2005), and the effects of the mutation on body morphology were found to resemble those seen in *C. elegans* *phy-1* (*dpy-18*)

mutants (Figure 2A, B, E and F in III). The *C. briggsae phy-1* mutants were shortened to the same extent as the latter, but no circumferential increase characteristic of a dumpy phenotype was observed (Figure 2B and F in III). The phenotype of the *C. briggsae phy-1* mutants is thus not as severe as that of its *C. elegans* counterpart, being small (Sma) rather than dumpy. To keep the nomenclature consistent, however, the *C. briggsae phy-1* mutants are also referred to as *C. briggsae dpy-18*.

Several gene rescue experiments were carried out on the *C. elegans* and *C. briggsae dpy-18* mutants. Genomic *phy-1* and *phy-2* rescue constructs that contained the wild-type genes from both species were coinjected with a visible cuticle collagen marker construct *dpy-7::gfp* and expression of the transgenes was verified by RT-PCR (Figure 1J in III). The *C. elegans dpy-18* phenotype was completely rescued by both *Ce-phy-1* and *Cb-phy-1* (Table 1 and Figure 1G in III), but surprisingly, the *C. briggsae dpy-18* phenotype was only partially rescued by *Cb-phy-1* and *Ce-phy-1* (Table 1 and Figure 1C in III). Injection of the *Ce-phy-2* rescue construct led to partial rescue of the *C. elegans dpy-18* phenotype but not of the *C. briggsae dpy-18* (Table 1 in III). The *Cb-phy-2* construct was unable to rescue either of the *dpy-18* mutants (Table 1 in III).

5.3.3 RNAi of *C. briggsae phy-1, phy-2 and pdi-2*

Simultaneous inactivation of the *C. elegans phy-1* and *phy-2* genes by RNAi or combined genetic mutations leads to a synthetic lethal phenotype (Friedman *et al.* 2000, Winter & Page 2000), which is explained by the loss of all three possible cuticle C-P4H forms (Myllyharju *et al.* 2002). To study the combined function of *C. briggsae phy-1* and *phy-2*, a number of RNAi experiments were carried out using wild-type *C. elegans* and *C. briggsae* and the *C. briggsae dpy-18* mutants. *Cb-phy-1* RNAi caused slight to medium dumpy and small phenotypes in wild-type *C. elegans* and *C. briggsae*, respectively (Table 2 in III), while *Cb-phy-1* RNAi did not increase the severity of the small phenotype in the *C. briggsae dpy-18* nematodes, indicating that these are true genetic nulls of the *phy-1* gene (Table 2 in III). *Cb-phy-2* RNAi produced severe dumpy and larval arrest with coiled larva phenotypes in the *C. briggsae dpy-18* mutants, whereas no visible phenotype was observed in the wild-type strains (Table 2 and Figure 3 in III). *Cb-pdi-2* RNAi caused severe dumpy and embryonic lethal phenotypes in both wild-type *Caenorhabditis* species (Table 2 in III).

5.3.4 Analysis of the collagen prolyl 4-hydroxylase forms generated from *C. briggsae PHY-1, PHY-2 and PDI-2* in insect cells and in vivo

Recombinant *C. briggsae* PHY-1 and PHY-2 were coexpressed singly or simultaneously with *C. briggsae* or *C. elegans* PDI-2 in insect cells. Similar experiments were carried out with recombinant *C. elegans* PHY-1 and PHY-2. The PHY-1 polypeptides from both species formed active dimers with PDI-2 (Figure 4 and Table 3 in III). Another feature in

common was that the PHY-1 and PHY-2 polypeptides from both species assembled into active mixed tetramers with PDI-2 (Figure 4 and Table 3 in III). The assembly of these enzyme forms was not dependent on the *Caenorhabditis* source of the PDI-2 subunit (Figure 4 and Table 3 in III) and a mixed tetramer was also formed when the PHY-1 and PHY-2 polypeptides were from the two different species. Interestingly, the results also showed that although the *C. briggsae* PHY-1 assembled into an active dimer with PDI-2 much more readily than the *C. elegans* PHY-1 when PHY-2 was not present (Figure 4 and Table 3 in III), assembly of the mixed tetramer rather than the PHY-1/PDI-2 dimer was more strongly favoured in *C. briggsae* than in *C. elegans* (Figure 4 in III). When all three recombinant polypeptides were coexpressed, both PHY-1/PDI-2 dimers and PHY-1/PHY-2/(PDI-2)₂ tetramers were formed from the *C. elegans* polypeptides, while the *C. briggsae* polypeptides assembled preferentially into the mixed tetramers and no PHY-1/PDI-2 dimers were formed (Figure 4 in III).

A distinct difference between the two species was that, unlike the *C. elegans* PHY-2, the *C. briggsae* PHY-2 readily formed an active complex with PDI-2 from both species (Figure 4 and Table 3 in III). The mobility of this novel complex in non-denaturing PAGE was between those of the mixed tetramer and the PHY-1/PDI-2 dimer (Figure 4 in III). Gel filtration analysis on a calibrated column showed that the active complex formed by *C. briggsae* PHY-2 and PDI-2 eluted in the same position as the mixed tetramer (Figure 5 in III), suggesting that it is a unique (PHY-2)₂/(PDI-2)₂ tetramer. The assembly of the *C. briggsae* PHY-2 into an active tetramer with PDI-2 was likewise independent of the *Caenorhabditis* origin of the PDI-2 (Figure 4 and Table 3 in III).

The *in vivo* assembly properties of the *C. briggsae* PHY-1, PHY-2 and PDI-2 polypeptides were studied by analyzing native extracts prepared from wild-type and *phy-1* mutant *C. briggsae* by non-denaturing PAGE. As with *C. elegans*, two C-P4H forms were present in the wild-type *C. briggsae*, i.e. PHY-1/PDI-2 dimers and PHY-1/PHY-2/(PDI-2)₂ tetramers (Figures 6 and 7 in III). As the insect cell expression experiments performed with the *C. briggsae* PHY-1, PHY-2 and PDI-2 polypeptides showed that formation of the mixed tetramer is strongly favoured over that of the PHY-1/PDI-2 dimer (Figure 4 in III), the *in vivo* data suggest a limitation in the *in vitro* availability of either the PHY-1 polypeptide or the PHY-2 polypeptide.

5.3.5 Analysis of the regions of the C. elegans and C. briggsae PHY polypeptides that determine their assembly properties

The data above showed that despite the high amino acid sequence identity between the *C. elegans* and *C. briggsae* PHY-1 and PHY-2 polypeptides, distinct differences exist in their assembly properties, with respect to both the assembly efficiency and specificity. The limited amino acid differences, mostly in the non-catalytic regions of the polypeptides (Figure 1 in III), could thus be expected to be responsible for the observed differences. A search for protein motifs in the PHY polypeptides suggested that they both have a TPR-like region between PHY-1 residues 134-237 and PHY-2 residues 144-226 (Figure 1 in III). These regions were predicted to contain two 34-residue TPR-motifs (Lamb *et al.* 1995), between PHY-1 residues 188-221 and 146-179 and PHY-2 residues

190-223 and 148-181 (Figure 1 in III). As TPR motifs mediate protein-protein and protein-peptide interactions and the assembly of multiprotein complexes (D'Andrea & Regan 2003), those in the PHY polypeptides were regarded as potential candidates for the determination of their assembly properties. The TPR motifs of the *C. elegans* and *C. briggsae* PHY homologues differ distinctly only in a few amino acid locations, two of which were selected for site-directed mutagenesis to change the *C. elegans* codon for the corresponding *C. briggsae* codon, and *vice versa* (Figure 1 in III), resulting in the generation of baculoviruses coding for the following mutant PHY polypeptides: CePHY-1-T220A, CbPHY-1-A220T, CePHY-2-W190S, CbPHY-2-S190W, CePHY-1-T220A/Q174K, CbPHY-1-A220T/K174Q, CePHY-2-W190S/V177N and CbPHY-2-S190W/N177V.

The mutant PHY polypeptides were coexpressed in insect cells together with *C. elegans* or *C. briggsae* PDI-2 and analyzed by non-denaturing PAGE and C-P4H activity assay. The single mutation in CePHY-1 did not enhance dimer formation, but the double mutant CePHY-1 formed active dimers with PDI-2 3-4-fold more efficiently (Table 4 and Figure 8 in III). The assembly efficiency of the double mutant CePHY-1 was still lower than that of the CbPHY-1, however, as reflected by the markedly lower level of C-P4H activity (Table 4 in III). The single mutation in the CbPHY-1 did not reduce its assembly efficiency with PDI-2 but rather increased it, while the double mutation led to about a 2-fold decrease in efficiency (Table 4 and Figure 8 in III). The assembly efficiency of the double mutant CbPHY-1 was still about 5-fold higher than that of CePHY-1, however (Table 4 in III).

The assembly properties of the single and double mutant CePHY-2 polypeptides were found to be altered towards those of the CbPHY-2. The double mutant CePHY-2 was able to form a homotetramer with PDI-2 (Figure 8 in III), which was also reflected by the about 8-fold increase in activity (Table 4 in III). This activity level was nevertheless still only 11% of that obtained with CbPHY-2 and PDI-2. In the case of the single mutant CePHY-2, the activity obtained was between those of the wild-type and double mutant CePHY-2/PDI-2 dimers. In contrast, the single and double mutations in the CbPHY-2 polypeptide did not reduce its assembly efficiency with PDI-2 (Table 4 and Figure 8 in III).

The above data suggest that the TPR motifs of the *C. elegans* and *C. briggsae* PHY polypeptides are involved in determining their assembly properties with PDI-2, but the differences in the two amino acids selected for site-directed mutagenesis do not fully explain the differences observed in their assembly properties. We therefore also studied the assembly properties of hybrid PHY-2 polypeptides in which the N-terminal 262 amino acids of the *C. briggsae* and *C. elegans* PHY-2 polypeptides had been exchanged. This N-terminal swapping was found to have a dramatic effect on the assembly properties. Replacement of this region of the *C. briggsae* PHY-2 with that of *C. elegans* resulted in a loss of the ability to form the $(\text{PHY-2})_2/(\text{PDI-2})_2$ tetramer (Figure 9 in III) and a 35-45-fold reduction in the C-P4H activity obtained (Table 4 in III), while substitution of the N-terminal part of the *C. elegans* PHY-2 for that of *C. briggsae* led to efficient formation of the $(\text{PHY-2})_2/(\text{PDI-2})_2$ tetramer (Figure 9 in III) and identical activity to that obtained with wild-type *C. briggsae* PHY-2 and PDI-2 (Table 4 in III).

6 Discussion

6.1 *C. reinhardtii* has multiple prolyl 4-hydroxylases, one of which is essential for proper cell wall assembly

Only three plant P4Hs from two species have been cloned to date, two from *A. thaliana* and one from *N. tabacum* (Hieta & Myllyharju 2002, Tiainen *et al.* 2005, Yuasa *et al.* 2005), while this thesis reports on the cloning of a fourth plant P4H, from the green alga *C. reinhardtii*. Although plants do not contain collagens, these hydroxylases are needed for the hydroxylation of various glycoproteins that are involved in a wide variety of tasks in plants, varying from structural roles to roles in plant growth and development. Previous analysis of the *A. thaliana* genome (Hieta *et al.* 2002, Tiainen *et al.* 2005) and the current analysis of the *C. reinhardtii* genome show that both species have a relatively large gene family coding for P4H-like polypeptides, the *A. thaliana* genome containing at least six such genes and the *C. reinhardtii* genome at least 10. The P4H families in both plant species are thus larger than the C-P4H families of vertebrates and *C. elegans*, which consist of only three and four isoenzymes, respectively (Bassuk *et al.* 1989, Helaakoski *et al.* 1989, Hopkinson *et al.* 1994, Veijola *et al.* 1994, Helaakoski *et al.* 1995, Annunen *et al.* 1997, Friedman *et al.* 2000, Hill *et al.* 2000, Winter & Page 2000, Riihimaa *et al.* 2002, Kukkola *et al.* 2003, Van Den Diepstraten *et al.* 2003, original article II of this thesis), but smaller than the *D. melanogaster* gene family of 19 members that codes for C-P4H α subunit-like polypeptides (Annuen *et al.* 1999, Abrams and Andrew 2002, Abrams *et al.* 2006). The requirement for such large gene families may be explained by differences in the expression patterns or substrate specificities of the various isoenzymes. The expression patterns of the vertebrate C-P4H isoenzymes differ from each other, for example, type I C-P4H being the main form in most cell types and tissues while the type II and III C-P4Hs have more specialized expression patterns (Annuen *et al.* 1998, Nissi *et al.* 2001, Kukkola *et al.* 2003, Van Den Diepstraten *et al.* 2003). Likewise, the genes coding for the various *C. elegans* C-P4H α subunit isoforms and the *D. melanogaster* C-P4H α subunit-like polypeptides display temporal and spatial differences in expression (Friedman *et al.* 2000, Hill *et al.* 2000, Winter & Page 2000, Abrams and Andrew 2002, Riihimaa *et al.* 2002, Abrams *et al.* 2006, original article II of this thesis). The identification of 10 *C. reinhardtii* P4Hs is a unique phenomenon, as this is the first time

that a single-cell organism has been shown to have a large number of P4H isoenzymes. Based on analysis of the as yet rather limited *C. reinhardtii* EST database, nine out of the ten P4H sequences seem to be present in the vegetative cells and one of them in the gametes. The nine P4Hs present in the vegetative cells may therefore either have major differences in their substrate specificities or may have redundant functions. The fact that Cr-P4H-1 turned out to be essential for proper cell wall assembly speaks for specialization with regard to substrate specificity. The expression patterns of the *A. thaliana* P4Hs have not been studied, but analysis of their catalytic properties has shown that they are also likely to have distinct *in vivo* substrates (Hieta & Myllyharju 2002, Tiainen *et al.* 2005).

The Cr-P4H-1 cloned and characterized here was shown to have two splicing variants: a shorter one, Cr-P4H-1A, which consists of 253 residues, while and a longer one, Cr-P4H-1B, with a C-terminal extension of 47 residues that has six cysteine residues characteristic of a ShK toxin-like domain. A similar domain is found in nine out of the 10 *C. reinhardtii* P4H sequences and also in At-P4H-2 (Tiainen *et al.* 2005). It is likely that the C-terminal extension in Cr-P4H-1B and in the other *C. reinhardtii* P4Hs does not function as a potassium channel toxin, however, since they lack the critical Lys-Tyr diad (Dauplais *et al.* 1997). In addition, two of the *C. reinhardtii* P4Hs also lack some of the six cysteines that are needed for disulphide bond formation and the proper folding and function of the ShK toxin. The C-terminal extension was not required for the P4H activity of Cr-P4H-1B, and its role in the *C. reinhardtii* P4H polypeptides has not yet been revealed.

Unlike animal C-P4Hs, the Cr-P4H-1A characterized here, the two P4Hs cloned from *A. thaliana* and those partially purified from other higher plants and *V. carteri* have been shown to exist as monomers that do not need a PDI/ β -subunit in order to be soluble or active (Tanaka *et al.* 1980, Bolwell *et al.* 1985, Kaska *et al.* 1988, Hieta & Myllyharju 2002, Tiainen *et al.* 2005). Although vertebrate C-P4H tetramers can be successfully produced as recombinant proteins in high amounts in insect cells, yeast and nowadays also in *E. coli* (Vuori *et al.* 1992a, Vuorela *et al.* 1997, Neubauer *et al.* 2006) and can be purified to homogeneity, all attempts to resolve the crystal structure of a vertebrate C-P4H tetramer have so far been unsuccessful. The structure of a monomeric HIF-P4H has been recently demonstrated (McDonough *et al.* 2006), but as C-P4Hs and HIF-P4Hs differ profoundly in their substrate specificity, the former hydroxylating prolines in repeating -X-Pro-Gly- sequences and the latter in an individual -Leu-X-X-Leu-Ala-Pro-sequence, distinct differences are likely to exist in their structures, even though they probably share a common jellyroll core fold. The structure of a monomeric plant P4H can thus be expected to shed light on the structural requirements for the hydroxylation of proline-rich proteins. Unlike the recombinant At-P4Hs, which are only poorly soluble in insect cells (Hieta & Myllyharju 2002, Tiainen *et al.* 2005), at least half of the recombinant Cr-P4H-1A was found to be soluble in insect cells, and especially in *E. coli*. Furthermore, it was found to be unusually easy to purify Cr-P4H-1A to homogeneity. These properties have led to recent successful crystallization of recombinant Cr-P4H-1A, which has been shown to have the typical 8-stranded jellyroll core fold, but also several unique structural properties relative to HIF-P4H that are likely to be required for efficient hydroxylation of proline-rich sequences (Koski *et al.*, unpublished results).

The plant P4Hs differ markedly from vertebrate C-P4Hs in that they efficiently hydroxylate poly(L-proline). None of the vertebrate C-P4Hs hydroxylates poly(L-proline), which instead acts as an effective competitive inhibitor for some of them. It was believed earlier that plant P4Hs do not hydroxylate collagen-like sequences at all or only very inefficiently, but this concept has changed upon the cloning and characterization of the recombinant *A. thaliana* (Hieta & Myllyharju 2002, Tiainen *et al.* 2005) and *C. reinhardtii* P4Hs, the efficiency of hydroxylation of the collagen-like peptide (Pro-Pro-Gly)₁₀ by At-P4H-1 actually being similar to those achieved by vertebrate C-P4Hs. Sequencing of (Pro-Pro-Gly)₁₀ after partial hydroxylation by At-P4H-1 and Cr-P4H-1A showed that these enzymes differ in their hydroxylation preference, the former resembling vertebrate C-P4Hs in that it acts preferentially on prolines that are in the Y position of the -X-Y-Gly- repeats (Hieta & Myllyharju 2002), while the latter preferentially hydroxylates prolines in the X position. Analysis of the hydroxylation of various proline-rich peptides by the recombinant At-P4Hs and Cr-P4H-1A has also shown that although they are all capable of hydroxylating poly(L-proline) they also have certain differences in their substrate sequence requirements.

C. reinhardtii HRGPs contain at least 3 kinds of proline-rich motif, consisting of strings of continuous proline residues, prolines alternating with other amino acids or repeats of a -Pro-Pro-Ser-Pro-X- sequence (Ferris *et al.* 2001). Cr-P4H-1A was found to hydroxylate effectively several synthetic peptides representing such motifs in domains 2 and 3 of the *C. reinhardtii* cell wall protein GP1, the sexual agglutinins Sag1 and Sad1 and a related protein A2, the function of which is currently unknown (Ferris *et al.* 2001, 2005). These findings prompted us to study the *in vivo* role of Cr-P4H-1 in the formation of the *C. reinhardtii* cell wall, which is a highly ordered structure that can be divided into seven structurally distinct layers based on electron microscopy analyses (Roberts *et al.* 1972, Goodenough & Heuser 1985, Adair & Snell 1990). The innermost and outermost layers, W1 and W7, respectively, are composed of anastomosing fibres that form a loose network, and the layers from W2 to W6 between them form a prominent central structure where the dense W2 and W6 layers are believed to have a crystalline structure and the W4 layer is composed of granules that are lined up and thus create a pearl-like structure. The layers W3 and W5 appear as empty spaces between the other central layers. The layer W6 is composed of three distinct HRGPs, GP1, GP2 and GP3 with 4-hydroxyproline contents of 32%, 15% and 6%, respectively (Goodenough *et al.* 1986, Adair & Snell 1990). Three-dimensional analyses of these three proteins have shown that their 4-hydroxyproline-rich regions adopt a poly(L-proline) type II helix conformation and form relatively long fibrous shafts that are believed to have an important role in the self-assembly of these proteins into distinct supramolecular structures (Goodenough & Heuser 1985, Goodenough *et al.* 1986). Furthermore, glycosylation is needed for the stabilization of the poly(L-proline) type II helix conformation of the shaft regions, and this modification also affects the interactive properties of GP proteins (Ferris *et al.* 2001). It has been suggested that two glycosylation codes may exist in higher plants, long, branching sugars being attached to non-contiguous 4-hydroxyproline residues via *O*-galactosyl linkages and short, unbranched sugars to contiguous 4-hydroxyproline residues via *O*-arabinosyl linkages (Shpak *et al.* 1999). The GP1 protein has characteristic 4Hyp-4Hyp-Ser-X-4Hyp- repeats and carries a complex mixture of sugar side chains, 37% of which are long and branching, while GP2 has contiguous 4-hydroxyproline residues and

a carbohydrate composition in which only 11% of the sugars are long and branching (Ferris *et al.* 2005). Knock-down of the Cr-P4H-1 by RNAi led to highly abnormal, typically non-multilayered cell walls formed of loose networks of fibres similar to those of layers W1 and W7, suggesting that Cr-P4H-1 activity is essential for the assembly of a normal cell wall. Since the cell walls of *C. reinhardtii* are formed in an ordered way, so that the long fibres of layers W1 and W7 emerge first and the rest of the layers assemble on this primary matrix later, it can be assumed that Cr-P4H-1 is needed for the assembly of the W2-W6 layers. In view of the *in vitro* hydroxylation properties of Cr-P4H-1, the lack of its activity most probably affected the hydroxylation and subsequent glycosylation of the three HRGPs GP1, GP2 and GP3, which are the major components of the W6 layer. The data clearly show that none of the nine other *C. reinhardtii* P4Hs can compensate for this lack of Cr-P4H-1 activity and are thus likely to have distinct substrate specificities and functions.

The plus and minus sexual agglutinins Sag1 and Sad1 of *C. reinhardtii* are needed for sexual adhesion between gametes of opposite mating types. These HRGPs are displayed on the flagellar membrane surfaces of the gametes and they have characteristic globular heads followed by a long, fibrous shaft with a poly(L-proline) type II conformation and a tail hook (Ferris *et al.* 2005). It has been suggested that differences in the amino acid composition of the proline-rich shafts of Sad1 and Sag1, and hence their 4-hydroxyproline patterns and concomitant glycosylation patterns, can affect shaft-shaft interactions and thus adhesion of the gametes. If Cr-P4H-1 turns out to be expressed in the gametes, it may also have an essential role in the mating of *C. reinhardtii*.

6.2 A novel member of the *C. elegans* prolyl 4-hydroxylase family

Analysis of the *C. elegans* genome indicates that its collagen gene family is likely to consist of over 170 genes coding for cuticle collagen polypeptides and three genes coding for basement membrane collagen polypeptides (for reviews, see Johnstone 2000, Page 2001, Page & Winter 2003, Myllyharju & Kivirikko 2004). The *C. elegans* genome contains five genes coding for putative C-P4H α subunits, three of them having been characterized to date (Veijola *et al.* 1994, Friedman *et al.* 2000, Hill *et al.* 2000, Winter & Page 2000, Riihimaa *et al.* 2002). We completed the characterization of this gene family here and showed that only one of the two remaining α subunit-like genes, namely Y43F8B.4, codes for a functional α subunit. *C. elegans* thus has four distinct α subunit isoforms, called PHY-1, PHY-2, PHY-3 and PHY-4. Three alternatively spliced forms of PHY-4 were found to be generated, although only one of them, PHY-4.1, functions as an α subunit. The *C. elegans* PHY polypeptides are much more versatile in their assembly properties (Figure 5) than the vertebrate α subunits, which are only capable of assembling into $\alpha_2\beta_2$ tetramers with the PDI/ β subunit, and only in such a way that the α subunits in the tetramer are identical. The *C. elegans* PHY-1 and PHY-2 polypeptides assemble mainly into a mixed PHY-1/PHY-2/(PDI-2)₂ tetramer, but they can also assemble into active PHY-1/PDI-2 and PHY-2/PDI-2 dimers, especially when formation of a mixed tetramer is impossible due to the lack of either PHY-1 or PHY-2 (Figure 5). The third *C. elegans* α subunit isoform, PHY-3, requires PDI-1 instead of PDI-2 for the formation of

an active C-P4H. Whether a complex is formed from PHY-3 and PDI-1, or whether PHY-3 is a functional C-P4H as such and requires PDI-1 only for correct folding, is as yet unknown (Fig. 5). The assembly properties of recombinant PHY-4.1 were in any case found to differ from those of the other *C. elegans* PHY isoforms, as it assembled into active (PHY-4.1)₂/(PDI-2)₂ tetramers as well as PHY-4.1/PDI-2 dimers in insect cells (Figure 5).

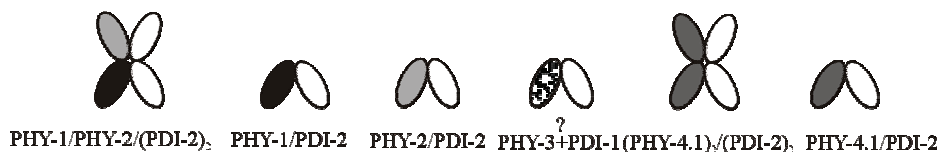


Fig. 5. Schematic representation of the C-P4H forms assembled from *C. elegans* PHY-1, PHY-2, PHY-3 and PHY-4.1 polypeptides.

The *C. elegans* C-P4H forms assembled from the PHY-1 and PHY-2 polypeptides have been shown to be involved in the hydroxylation of cuticle collagens (Friedman *et al.* 2000, Hill *et al.* 2000, Winter & Page 2000, Myllyharju *et al.* 2002). This function is supported by their expression in hypodermal cells, the level of which peaks at times of maximal cuticle collagen synthesis, coinciding with moulting. PHY-3 has a more restricted spatial expression, as it is only found in the spermatheca, and it probably has a role in the synthesis of egg shell collagens (Riihimaa *et al.* 2002). Expression of PHY-4.1 was also spatially restricted, as it was only found in the pharynx, at the boundary between the pharynx and the gut and in the excretory duct.

The substrate specificity of the active tetramers and dimers formed from PHY-4.1 and PDI-2 indicates that a separate class of P4Hs is to be found in *C. elegans* that is distinct from the C-P4Hs and HIF-P4Hs. A collagen-like sequence was hydroxylated relatively ineffectively by the (PHY-4.1)₂/(PDI-2)₂ tetramers and PHY-4.1/PDI-2 dimers, while they hydroxylated a non-collagenous proline-rich synthetic peptide (Pro-Glu-Pro-Pro-Ala)₅ more effectively and were also capable of hydroxylating poly(L-proline). With respect to their substrate requirements, these novel *C. elegans* P4H forms assembled from PHY-4.1 and PDI-2 thus resemble plant P4Hs rather than the animal C-P4Hs and HIF-P4Hs.

The *C. elegans* pharyngeal gland cells are not known to express collagens and thus the unique spatial expression of PHY-4.1 also suggests that it probably hydroxylates some other proline-rich *C. elegans* proteins rather than collagens. Interestingly, expression of two of the 19 *D. melanogaster* α subunit-like polypeptides, PH4 α SG1 and PH4 α SG2, is restricted to the salivary glands, which do not express any of the three known *D. melanogaster* collagen polypeptides (Abrams & Andrew 2002, Abrams *et al.* 2006). A large number of uncharacterized genes that encode proline-rich repeats are expressed in the fly salivary glands, however, and it has been hypothesized that some of these may be secreted or transmembrane proteins that act as substrates for PH4 α SG1 and PH4 α SG2 (Abrams *et al.* 2006). Furthermore, PH4 α SG1 and PH4 α SG2 were found to be necessary for the maintenance of the proper structure and secretory function of the salivary gland lumen (Abrams *et al.* 2006). The salivary glands of *D. melanogaster* resemble morphologically the nematode pharyngeal gland cells and excretory duct, and it is thus

possible that PHY-4.1 has a corresponding role in *C. elegans*. It will therefore be of interest to study the morphology of the pharynx and the secretory duct in detail once a deletion mutant of *phy-4* is available. A mutant allele *tm1040* of the Y43F8B.4 gene has been generated by the *C. elegans* Gene Knockout Project using random mutagenesis, but these mutant nematodes have been reported to be either lethal or sterile, which does not agree with the wild-type appearance of the PHY-4.1 knock-down nematodes in our RNAi experiments. The *phy-4* gene is located at the intron of another putative gene, Y43F8B.3, that shows homology with the Kunitz/Bovine pancreatic trypsin inhibitor domain. It is not likely that the lethal or sterile phenotype reported in *tm1040* mutant nematodes is caused by concomitant interference from the Y43F8B.3 gene, since RNAi experiments with this gene did not result in any obvious phenotype (Kamath *et al.* 2003). It is therefore likely that the lethality or sterility of the *tm1040* mutant nematodes will have been caused by some other unrelated genetic defects generated during the trimethylpsoralen-UV mutagenesis treatment. Outcrossing is therefore required before any detailed analysis of the *tm1040* mutant nematodes can be commenced.

6.3 Assembly of cuticle collagen prolyl 4-hydroxylases in *C. elegans* and *C. briggsae*

C. elegans PHY-1, PHY-2 and PDI-2 have previously been shown to assemble into unique C-P4H forms that are responsible for the hydroxylation of cuticle collagens (Friedman *et al.* 2000, Winter & Page 2000, Myllyharju *et al.* 2002). Analysis of the *C. briggsae* genome indicated the presence of genes encoding highly homologous C-P4H subunits. Despite the high amino acid sequence identity, certain differences were found in the assembly of *C. briggsae* cuticle C-P4Hs relative to those of *C. elegans*, however. The *C. briggsae* cuticle C-P4H subunits also assemble into active PHY-1/PHY-2/(PDI-2)₂ tetramers and PHY-1/PDI-2 dimers, but instead of a PHY-2/PDI-2 dimer an active (PHY-2)₂/(PDI-2)₂ tetramer is formed. The molecular composition of the (PHY-2)₂/(PDI-2)₂ tetramer found in *C. briggsae* represents the consensus subunit composition of all other known animal C-P4Hs, mixed tetramers and dimers having been so far identified only in the *Caenorhabditis* species.

C. elegans and *C. briggsae* nematodes that lack a functional *phy-1* or *phy-2* gene are viable, but combined inactivation of these genes leads to a synthetic lethal phenotype due to a lack of cuticle collagen synthesis. The absence of PHY-2 does not lead to any obvious phenotypic abnormalities in either nematode species. Analysis of *C. elegans phy-2* null deletion mutants has shown that they can compensate for a lack of the PHY-1/PHY-2/(PDI-2)₂ tetramers by increasing the assembly of PHY-1/PDI-2 dimers (Myllyharju *et al.* 2002). The *C. elegans phy-2* mutants are thus able to retain 54-57% of the wild-type C-P4H activity (Myllyharju *et al.* 2002) and 64% of the cuticle 4-hydroxyproline content (Friedman *et al.* 2000), these levels being enough to maintain the wild-type phenotype. *C. briggsae* is likely to have a same kind of compensation mechanism, since its PHY-1 was also found to readily form an active dimer with PDI-2.

The absence of PHY-1 leads to abnormal cuticle synthesis that results in a deformed body shape in both *Caenorhabditis* species. The deformations are not identical, however,

the *C. briggsae phy-1* null mutants having a small (Sma) body size, while the corresponding *C. elegans* mutants have a small, fat (Dpy) body. The cuticle is composed of two distinct regions, derived from lateral hypodermal seam cells and the dorso-ventral hypodermis (Sulston & Horvitz 1977). These two regions are morphologically and functionally distinct and it has been shown that mutations in different collagen genes can change one or both of these areas thus creating phenotypes that are defined by region (Thein *et al.* 2003). The Sma phenotype is less severe than the Dpy phenotype, as it affects only the lateral cuticle structures while the latter affects both the lateral and dorso-ventral cuticle regions. This difference in phenotypic outcome between the *C. elegans* and *C. briggsae phy-1* mutants can probably be explained by the distinct differences observed between the assembly properties of the recombinant PHY-2 polypeptides of the two species. It is likely that the absence of the PHY-1 polypeptide is compensated for more efficiently in *C. briggsae* by the active (PHY-2)₂/(PDI-2)₂ tetramers that are readily formed. In contrast, the assembly of *C. elegans* PHY-2 into an active PHY-2/PDI-2 dimer is likely to be much less effective, as is also reflected in the marked reduction of 97.5-99.3% in the C-P4H activity of the *C. elegans phy-1* mutants relative to the wild type (Myllyharju *et al.* 2002).

The compensatory potential of the *C. briggsae* and *C. elegans phy-1* and *phy-2* genes was studied by means of rescue experiments conducted on the *phy-1 (dpy-18)* mutants of both species. The *phy-1* transgenes from both species were capable of both intra-species and inter-species rescue, the phenotypic rescue of the *C. elegans phy-1* mutants being complete, while that of the *C. briggsae phy-1* mutants was partial. This clearly indicates that the PHY-1 polypeptides of both species share structural and functional properties of the same kind. The most likely explanations for the only partial rescue of the *C. briggsae phy-1* mutants would be that the *Ce-phy-1* transgene construct was toxic for levels, and the fact that the *Cb-phy-1* transgene was flanked by the *C. elegans phy-1* 5' and 3' regulatory regions, which may affect its function in *C. briggsae*. A highly surprising finding was the partial rescue of the *C. elegans phy-1* mutants by the *Ce-phy-2* transgene, which has not been observed in earlier studies. This is likely to be explained by an increase in the amount of PHY-2/PDI-2 dimer facilitated by additional copies of a functional *phy-2* gene. The reason why such partial rescue has not previously been obtained may be that genomic rescue constructs were used in the current study rather than cDNA constructs. The lack of rescue of the *C. elegans* and *C. briggsae phy-1* mutants by inter-species *phy-2* transgenes, and in the case of the latter mutant also an intra-species *phy-2* transgene, may not necessarily rule out any compensatory potential, as it may also be caused by technical limitations. However, the lack of compensation in these experiments may also reflect true differences in the substrate specificities and/or assembly efficiencies of the C-P4H forms that have PHY-2 as their single catalytic subunit.

The different assembly properties of the highly homologous *C. elegans* and *C. briggsae* PHY-2 polypeptides, with PDI-2 facilitated identification of critical PHY-2 regions that determine their assembly into either dimers or tetramers. Previous studies with *C. elegans* PHY-1 and PHY-2 have shown that residues Gln121-Ala271 and Asp1-Leu122, respectively, are critical for the assembly of the mixed tetramer PHY-1/PHY-2/(PDI-2)₂ (Myllyharju *et al.* 2002), indicating that the N-terminal parts of the polypeptides have a crucial role in this assembly. Two TPR-like motifs were predicted to

be present in the N-terminal halves of the PHY-2 polypeptides, between residues 148-223. As these motifs are known to be important in many protein-protein interactions and in the assembly of multiprotein complexes (D'Andrea & Regan 2003), they seemed to be potential candidates for the determination of the assembly properties of the PHY-2 polypeptides. The *C. elegans* and *C. briggsae* PHY-2 TPR motifs differ from each other in only a few amino acids, and simultaneous mutation of two of them to those of the other species altered their assembly properties in the direction of those of the other species. A complete reversal was obtained, however, only when the entire N-terminal halves of the polypeptides were exchanged. This indicates that, although TPR motifs are involved in the determination of the assembly properties of the PHY-2 polypeptides with PDI-2, they are not the sole determinants, but other adjacent residues in the N-terminal halves also play a crucial role.

References

- Abrams EW & Andrew DJ (2002) Prolyl 4-hydroxylase α -related proteins in *Drosophila melanogaster*: tissue-specific embryonic expression of the 99F8-9 cluster. *Mech Dev* 112: 165-171.
- Abrams EW, Mihoulides WK & Andrew DJ (2006) Fork head and Sage maintain a uniform and patent salivary gland lumen through regulation of two downstream target genes, PH4 α SG1 and PH4 α SG2. *Development* 133: 3517-3527.
- Ackley BD, Crew JR, Elamaa H, Pihlajaniemi T, Kuo CJ & Kramer JM (2001) The NC1/endostatin domain of *Caenorhabditis elegans* type XVIII collagen affects cell migration and axon guidance. *J Cell Biol* 152: 1219-1232.
- Ackley BD, Kang SH, Crew JR, Suh C, Jin Y & Kramer JM (2003) The basement membrane components nidogen and type XVIII collagen regulate organization of neuromuscular junctions in *Caenorhabditis elegans*. *J Neurosci* 23: 3577-3587.
- Adachi T, Tomita M & Yoshizato K (2005) Synthesis of prolyl 4-hydroxylase α subunit and type IV collagen in hemocytic granular cells of silkworm, *Bombyx mori*: Involvement of type IV collagen in self-defense reaction and metamorphosis. *Matrix Biol* 24: 136-154.
- Adair WS & Apt KE (1990) Cell wall regeneration in *Chlamydomonas*: accumulation of mRNAs encoding cell wall hydroxyproline-rich glycoproteins. *Proc Natl Acad Sci USA* 87: 7355-7359.
- Adair WS & Snell W (1990) The *Chlamydomonas reinhardtii* cell wall: structure, biochemistry, and molecular biology. In: Adair WS & Mecham R (eds) *Organization and Assembly of Plant and Animal Extracellular Matrix*. Academic Press, Inc, San Diego, 15-84.
- Albertson DG & Thomson JN (1976) The pharynx of *Caenorhabditis elegans*. *Philos Trans R Soc Lond B Biol Sci* 275: 299-325.
- 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 the 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.
- Aravind L & Koonin EV (2001) The DNA-repair protein AlkB, EGL-9, and leprecan define new families of 2-oxoglutarate- and iron-dependent dioxygenases. *Genome Biol* 2: research0007.1-0007.8.

- Barr MM (2003) Super models. *Physiol Genomics* 13: 15-24.
- Bassuk JA, Kao WW, Herzer P, Kedersha NL, Seyer J, DeMartino JA, Daugherty BL, Mark GE 3rd & Berg RA (1989) Prolyl 4-hydroxylase: molecular cloning and the primary structure of the α subunit from chicken embryo. *Proc Natl Acad Sci USA* 86: 7382-7386.
- Bell A, Monaghan P & Page AP (2006) Peptidyl-prolyl cis-trans isomerases (immunophilins) and their roles in parasite biochemistry, host-parasite interaction and antiparasitic drug action. *Int J Parasitol* 36: 261-276.
- 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* 15: 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.
- Bergmann DC, Crew JR, Kramer JM & Wood WB (1998) Cuticle chirality and body handedness in *Caenorhabditis elegans*. *Dev Genet* 23: 164-174.
- Bessereau JL, Wright A, Williams DC, Schuske K, Davis MW & Jorgensen EM (2001) Mobilization of a *Drosophila* transposon in the *Caenorhabditis elegans* germ line. *Nature* 413: 70-74.
- Bieri T, Blasiar D, Ozersky P, Antoshechkin I, Bastiani C, Canaran P, Chan J, Chen N, Chen WJ, Davis P, Fiedler TJ, Girard L, Han M, Harris TW, Kishore R, Lee R, McKay S, Müller H-M, Nakamura C, Petcherski A, Rangarajan A, Rogers A, Schindelman G, Schwarz EM, Spooner W, Tuli MA, Van Auken K, Wang D, Wang X, Williams G, Durbin R, Stein LD, Sternberg PW & Spieth J (2007) Wormbase: new content and better access. *Nucl Acids Res* 35: D506-D510.
- 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.
- Cassab GI (1998) Plant cell wall proteins. *Ann Rev Plant Phys Plant Mol Biol* 49: 281-309.
- C. elegans* Genome Consortium (1998) Genome sequence of the nematode *C. elegans*: a platform for investigative biology. *Science* 282: 2012-2018.
- Chareyre P, Besson MT, Fourche J & Bosquet G (1996) Identification of a *Bombyx* collagenous protein with multiple short domains of Gly-Xaa-Yaa repeats: cDNA characterization and regulation of expression. *Insect Biochem Mol Biol* 26: 677-685.
- Chartier A, Zaffran S, Astier M, Semeriva M & Gratecos D (2002) Pericardin, a *Drosophila* type IV collagen-like protein is involved in the morphogenesis and maintenance of the heart epithelium during dorsal ectoderm closure. *Development* 129: 3241-3253.
- Clark J & Ding S (2006) Generation of RNAi libraries for high-throughput screens. *J Biomed Biotechnol* 2006: 1-7.
- Cleary EG & Gibson MA (1996) Elastic tissue, elastin and elastin associated microfibrils. In: Comper WD (ed) *Extracellular Matrix, Molecular Components and Interactions*. Volume 2. Harwood Academic Publishers GmbH, The Netherlands, 95-140.
- Clifton IJ, McDonough Ma, Ehrismann D, Kershaw NJ, Granatino N & Schofield CJ (2006) Structural studies on 2-oxoglutarate oxygenases and related double-stranded β -helix fold proteins. *J Inorg Biochem* 100: 644-669.
- Culetto E & Sattelle DB (2000) A role for *Caenorhabditis elegans* in understanding the function and interactions of human diseases. *Hum Mol Gen* 9: 869-877.
- D'Andrea LD & Regan L (2003) TPR proteins: the versatile helix. *Trends Biochem Sci* 28: 655-662.

- Dann CE 3rd, Bruick RK & Deisenhofer J (2002) Structure of factor-inhibiting hypoxia-inducible factor 1: An asparaginyl hydroxylase involved in the hypoxic response pathway. *Proc Natl Acad Sci USA* 99: 15351-15356.
- Dauplais M, Lecoq A, Song J, Cotton J, Jamin N, Gilquin B, Roumestand C, Vita C, de Medeiros CL, Rowan EG, Harvey AL & Menez A (1997) On the convergent evolution of animal toxins. Conservation of a diad of functional residues in potassium channel-blocking toxins with unrelated structures. *J Biol Chem* 272: 4302-4309.
- 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 ¹H, ¹³C and ¹⁵N resonances of the α' domain of protein disulfide isomerase. *J Biomol NMR*: 195-196.
- Edens WA, Sharling L, Cheng GJ, Shapira R, Kinkade JM, Lee T, Edens HA, Tang XX, Sullards C, Flaherty DB, Benian GM & Lambeth JD (2001) Tyrosine cross-linking of extracellular matrix is catalyzed by Duox, a multidomain oxidase/oxidoreductase with homology to the phagocyte oxidase subunit gp91 *phox*. *J Cell Biol* 154: 879-891.
- Elbashir SM, Martinez J, Patkaniowska A, Lendeckel W & Tuschl T (2001) Functional anatomy of siRNA for mediating efficient RNAi in *Drosophila melanogaster* embryo lysate. *EMBO J* 20: 6877-6888.
- Elkins JM, Hewitson KS, McNeill LA, Seibel JF, Schlemminger I, Pugh CW & Ratcliffe PJ (2003) Structure of factor-inhibiting hypoxia-inducible factor (HIF) reveals mechanism of oxidative modification of HIF-1 α . *J Biol Chem* 278: 1802-1806.
- Elkins JM, Ryle MJ, Clifton IJ, Dunning Hotopp JC, Lloyd JS, Burzlaff NI, Baldwin JE, Hausinger RP & Roach PL (2002) X-ray crystal structure of *Escherichia coli* taurine/ α -ketoglutarate dioxygenase complexed to ferrous iron and substrates. *Biochemistry* 23: 5185-5192.
- Ellgaard L & Ruddock LW (2005) The human protein disulphide isomerase family: substrate interactions and functional properties. *EMBO Rep* 6: 28-32.
- Epstein AC, Gleadle JM, McNeill LA, Hewitson KS, O'Rourke J, Mole DR, Mukherji M, Metzger 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.
- Eschenlauer SCP & Page AP (2003) The *Caenorhabditis elegans* ERp60 homolog protein disulfide isomerase-3 has disulfide isomerase and transglutaminase-like cross-linking activity and is involved in the maintenance of body morphology. *J Biol Chem* 278: 4227-4237.
- Ferris PJ, Waffenschmidt S, Umen JG, Lin H, Lee JH, Ishida K, Kubo T, Lau J & Goodenough UW (2005) Plus and minus sexual agglutinins from *Chlamydomonas reinhardtii*. *Plant Cell* 17: 597-615.
- Ferris PJ, Woessner JP, Waffenschmidt S, Kilz S, Drees J & Goodenough UW (2001) Glycosylated polyproline II rods with kinks as a structural motif in plant hydroxyproline-rich glycoproteins. *Biochemistry* 40: 2978-2987.
- 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.

- Frand AR, Russel S & Ruvkun G (2005) Functional genomic analysis of *C. elegans* molting. *PLoS Biol* 3: e312.
- Franklin SE & Mayfield SP (2004) Prospects for molecular farming in the green alga *Chlamydomonas reinhardtii*. *Plant Biotech* 7: 159-165.
- Fraser AG, Kamath RS, Zipperlen P, Martinez-Campos M, Sohrmann M & Ahringer J (2000) Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference. *Nature* 408: 325-330.
- Freedman RB, Klappa P & Ruddock LW (2002) Protein disulfide isomerases exploit synergy between catalytic and specific binding domains. *EMBO Rep* 3: 136-140.
- 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.
- Fuentes C & VanWinkle-Swift K (2003) Isolation and characterization of a cell wall-defective mutant of *Chlamydomonas monoica* (Chlorophyta). *J Phycol* 39: 1261-1267.
- Gallo M, Mah AK, Johnsen RC, Rose AM & Baillie DL (2006) *Caenorhabditis elegans dpy-14*: an essential collagen gene with unique expression profile and physiological roles in early development. *Mol Genet Genomics* 275: 527-539.
- Goodenough UW, Gebhart B, Mecham RP & Heuser JE (1986) Crystals of the *Chlamydomonas reinhardtii* cell wall: polymerization, depolymerization, and purification of glycoprotein monomers. *J Cell Biol* 103:405-417.
- Goodenough UW & Heuser JE (1985) The *Chlamydomonas* cell wall and its constituent glycoproteins analyzed by the quick-freeze, deep-etch technique. *J Cell Biol* 101: 1550-1568.
- Graham PL, Johnson JJ, Wang S, Sibley MH, Gupta MC & Kramer JM (1997) Type IV collagen is detectable in most, but not all, basement membranes of *Caenorhabditis elegans* and assembles on tissues that do not express it. *J Cell Biol* 137: 1171-1183.
- Griesbeck C, Kobl I & Heitzer M (2006) *Chlamydomonas reinhardtii*: a protein expression system for pharmaceutical and biotechnological proteins. *Mol Biotechnol* 34: 213-223.
- Grishok A (2005) RNAi mechanism in *Caenorhabditis elegans*. *FEBS Lett* 579: 5932-5939.
- Grossman AR, Harris EE, Hauser C, Lefebvre PA, Matrinez D, Rokhsar D, Shrager J, Silflow CD, Stern D, Vallon O & Zhang Z (2003) *Chlamydomonas reinhardtii* at the crossroads of genomics. *Eukaryot Cell* 2: 1137-1150.
- Grossman AR, Lohr M & Im CS (2004) *Chlamydomonas reinhardtii* in the landscape of pigments. *Annu Rev Genet* 38: 119-173.
- Guo X & Kramer JM (1989) The two *Caenorhabditis elegans* basement membrane (type IV) collagen genes are located on separate chromosomes. *J Biol Chem* 264: 17575-17582.
- Guo X, 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 $\alpha 1(IV)$ collagen mutations in *Caenorhabditis elegans* and the effects of $\alpha 1$ and $\alpha 2(IV)$ mutations on type IV collagen distribution. *J Cell Biol* 137: 1185-1196.
- Gupta BP & Sternberg PW (2003) The draft genome sequence of the nematode *Caenorhabditis briggsae*, a companion to *C. elegans*. *Genome Biol* 4: 238.
- Gönczy P, Echeverri C, Oegema J, Coulson AR, Jones SJM, Copley RR, Duperon J, Oegema J, Brehm M, Cassin E, Hannak E, Kirkham M, Pichler SC, Flohrs K, Goessen A, Leidel S, Alleaume AM, Martin C, Ozlu N, Bork P & Hyman AA (2000) Functional genomic analysis of cell division in *C. elegans* using RNAi of genes on chromosome III. *Nature* 408: 331-336.
- Hanuske-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.

- Hanauske-Abel HM & Popowicz AM (2003) The HAG mechanism: a molecular rationale for the therapeutic application of iron chelators in human diseases involving the 2-oxoacid utilizing dioxygenases. *Curr Med Chem* 10: 1005-1019.
- Hariharan IK & Haber DA (2003) Yeast, flies, worms, and fish in the study of human disease. *N Engl J Med* 348: 2457-2463.
- Harris EH (2001) *Chlamydomonas* as a model organism. *Annu Rev Plant Physiol Plant Mol Biol* 52: 363-406.
- Hausinger RP (2004) FeII/ α -ketoglutarate-dependent hydroxylases and related enzymes. *Crit Rev Biochem Mol Biol* 39: 21-68.
- Helaakoski T, Annunen P, Vuori K, McNeil 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 L, Taillon-Miller P, Kivirikko KI & Pihlajaniemi T (1994) Structure and expression of the human gene for the α subunit of prolyl 4-hydroxylase. *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.
- 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.
- Holmgren SK, Bretscher LE, Taylor KM & Raines RT (1999) A hyperstable collagen mimic. *Chem Biol* 6: 63-70.
- Holster T, Pakkanen O, Soininen R, Sormunen R, Nokelainen M, Kivirikko KI & Myllyharju J (2007) Loss of assembly of the main basement membrane collagen, type IV, but not fibril-forming collagens and embryonic death in collagen prolyl 4-hydroxylase I null mice. *J Biol Chem* 282: 2512-2519.
- Hopkinson I, Smith SA, Donne A, Gregory H, Franklin TJ, Grant ME & Rosamond J (1994) The complete cDNA derived sequence of the rat prolyl 4-hydroxylase α subunit. *Gene* 149: 391-392.
- Hynes RO & Zhao Q (2000) The evolution of cell adhesion. *J Cell Biol* 150: F89-96.
- 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* 15: 13459-13464.
- Ivan M, Kondo K, Yang H, Kim W, Valiando J, Ohh M, Salic A, Asara JM, Lane WS & Kaelin WG Jr (2001) HIF α targeted for VHL-mediated destruction by proline hydroxylation: implications for O₂ sensing. *Science* 292: 464-468.
- Jaakkola P, Mole DR, Tian Y-M, Wilson MI, Gielbert J, Gaskell SJ, 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 O₂-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.

- John DC & Bulleid NJ (1994) Prolyl 4-hydroxylase: defective assembly of alpha-subunit mutants indicates that assembled alpha-subunits are intramolecularly disulfide bonded. *Biochemistry* 33: 14018-14025.
- Johnstone IL (2000) Cuticle collagen genes. Expression in *Caenorhabditis elegans*. *Trends Genet* 16: 21-27.
- Johnstone IL, Shafi Y & Barry JD (1992) Molecular analysis of mutations in the *Caenorhabditis elegans* collagen gene *dpy-7*. *EMBO J* 11: 3857-3863.
- Jones AK, Buckingham SD & Sattelle DB (2005) Chemistry-to-gene screens in *Caenorhabditis elegans*. *Nat Rev Drug Discov*: 321-330.
- Kaelin WG Jr (2005) Proline hydroxylation and gene expression. *Ann Rev Biochem* 74: 115-128.
- Kaletta T & Hengartner MO (2006) Finding functions in novel targets: *C. elegans* as a model organism. *Nat Rev Drug Discov* 21: 387-399.
- Kamath RS, Fraser AG, Dong Y, Poulin G, Durbin R, Gotta M, Kanapin A, Le Bot N, Moreno S, Sohrmann M, Welchman DP, Zipperlen P & Ahringer J (2003) Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* 421: 231-237.
- Kaska DD, Günzler V, Kivirikko KI & Myllylä R (1987) Characterization of a low-relative-molecular-mass prolyl 4-hydroxylase from the green alga *Chlamydomonas reinhardtii*. *Biochem J* 241: 483-490.
- Kaska DD, Myllylä R, Günzler V, Gibor A & Kivirikko KI (1988) Prolyl 4-hydroxylase from *Volvox carteri*. *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 $^{13}\text{C}/^{15}\text{N}$ 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.
- Kemmink J, Dijkstra K, Mariani M, Scheek RM, Penka E, Nilges M & Darby NJ (1999) The structure in solution of the b domain of protein disulfide isomerase. *J Biomol NMR* 13: 357-368.
- Kieliszewski MJ & Lampion DTA (1994) Extensin: repetitive motifs, functional sites, post-translational codes, and phylogeny. *Plant J* 5: 157-172.
- Kieliszewski MJ & Shpak E (2001) Synthetic genes for the elucidation of glycosylation codes for arabinogalactan-proteins and other hydroxyproline-rich glycoproteins. *Cell Mol Life Sci* 58: 1386-1398.
- Kim TH, Hwang SB, Jeong P-Y, Lee J & Cho JW (2005) Requirement of tyrosylprotein sulfotransferase-A for proper cuticle formation in the nematode *C. elegans*. *FEBS Lett* 579: 53-58.
- Kivirikko KI & Myllyharju J (1998) Prolyl 4-hydroxylases and their protein disulfide isomerase subunit. *Matrix Biol* 16: 357-368.
- 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 JC (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.
- Knox JP (1995) The extracellular matrix in higher plants. 4. Developmentally regulated proteoglycans and glycoproteins of the plant cell surface. *FASEB J* 9: 1004-1012.
- Ko FCF & Chow KL (2002) A novel thioredoxin-like protein encoded by the *C. elegans dpy-11* gene is required for body and sensory organ morphogenesis. *Development* 129: 1185-1194.

- Koivunen P, Pirneskoski A, Karvonen P, Ljung J, Helaakoski T, Notbohm H & Kivirikko KI (1999) The acidic C-terminal domain of protein disulfide isomerase is not critical for the enzyme subunit function or for the chaperone or disulfide isomerase activities of the polypeptide. *EMBO J* 18: 65-74.
- Koivunen P, Salo KEH, Myllyharju J & Ruddock LW (2005) Three binding sites in protein-disulfide isomerase cooperate in collagen prolyl 4-hydroxylase tetramer assembly. *J Biol Chem* 280: 5227-5235.
- Kramer JM (1994) Structures and functions of collagens in *Caenorhabditis elegans*. *FASEB J* 8: 329-336.
- Kramer JM, French RP, Park EC & Johnson JJ (1990) The *Caenorhabditis elegans rol-6* gene, which interacts with the *sqt-1* collagen gene to determine organismal morphology, encodes a collagen. *Mol Cell Biol* 10: 2081-2089.
- Kramer JM, Johnson JJ, Edgar RS, Basaj C & Roberts S (1988) The *sqt-1* gene of *C. elegans* encodes a collagen critical for organismal morphogenesis. *Cell* 55: 555-565.
- 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* 278: 47685-47693.
- Kukkola L, Koivunen P, Pakkanen O, Page AP & Myllyharju J (2004) Collagen prolyl 4-hydroxylase tetramers and dimers show identical decreases in K_m values for peptide substrates with increasing chain length: mutation of one of the two catalytic sites in the tetramer inactivates the enzyme by more than half. *J Biol Chem* 279: 18656-18661.
- Lamb JR, Tudendreich S & Hieter P (1995) Tetratricopeptide repeat interactions: to TPR or not to TPR? *Trends Biochem Sci* 20: 257-259.
- 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.
- Lappe JC & Roy R (2006) New developmental insights from high-throughput biological analysis in *Caenorhabditis elegans*. *Clin Genet* 69: 306-314.
- 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.
- Lefebvre PA & Silflow CD (1999) *Chlamydomonas*: The cell and its genomics. *Genetics* 151: 9-14.
- Levy AD, Yang J & Kramer JM (1993) Molecular and genetic analyses of the *Caenorhabditis elegans dpy-2* and *dpy-10* collagen genes: a variety of molecular alterations affect organismal morphology. *Mol Biol Cell* 4: 803-817.
- Lunstrum GP, Bächinger HP, Fessler LI, Duncan KG, Nelson RE & Fessler JH (1988) *Drosophila* basement membrane procollagen IV. I. Protein characterization and distribution. *J Biol Chem* 263: 18318-18327.
- Maeda I, Kohara Y, Yamamoto M & Sugimoto A (2001) Large-scale analysis of gene function in *Caenorhabditis elegans* by high-throughput RNAi. *Curr Biol* 11: 171-176.
- McDonough MA, Li V, Flisman E, Chowdhury R, Mohr C, Liénard BMR, Zondlo J, Oldham NJ, Clifton IJ, Lewis J, McNeill LA, Kurzeja RJM, Hewitson KS, Yang E, Jordan S, Syed RS & Schofield CJ (2006) Cellular oxygen sensing: Crystal structure of hypoxia-inducible factor prolyl hydroxylase (PHD2). *Proc Natl Acad Sci USA* 103: 9814-9819.
- McGinnis K, Ku GM, VanDusen VJ, Fu J, Garsky V, Stern AM & Friedman PA (1996) Site-directed mutagenesis of residues in a conserved region of bovine aspartyl (asparaginyl) β -hydroxylase: evidence that histidine 675 has a role in binding Fe^{2+} . *Biochemistry* 35: 3957-3962.
- 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.
- Mizohata E, Matsumura H, Okano Y, Kumei M, Takuma H, Onodera J, Kato K, Shibata N, Inoue T, Yokota A & Kai Y (2002) Crystal structure of activated ribulose-1,5-bisphosphate carboxylase/oxygenase from green alga *Chlamydomonas reinhardtii* complexed with 2-carboxyarabinitol-1,5-bisphosphate. *J Mol Biol* 316: 679-691.
- Myllyharju J (2003) Prolyl 4-hydroxylases, the key enzymes of collagen biosynthesis. *Matrix Biol* 22: 15-24.
- Myllyharju J (2005) Intracellular post-translational modifications of collagens. In Brinckmann J, Notbohm H & Müller PK (eds) *Topics in Current Chemistry*. Springer-Verlag, 247:115-147.
- 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 (2004) Collagens, modifying enzymes and their mutations in humans, flies and worms. *Trends Genet* 20: 33-43.
- 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.
- 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.
- Natsuka S, Takubo R, Seki R & Ikura K (2001) Molecular cloning and expression of *Caenorhabditis elegans* ERp57-homologue with transglutaminase activity. *J Biol Chem* 276: 731-735.
- Nelson FK, Albert PS & Riddle DL (1983) Fine structure of the *Caenorhabditis elegans* secretory-excretory system. *J Ultrastruct Res* 82:156-171.
- Neubauer A, Neubauer P & Myllyharju J (2005) High-level production of human collagen prolyl 4-hydroxylase in *Escherichia coli*. *Matrix Biol* 24: 59-68.
- Nickelsen J & Kuck U (2000) The unicellular green alga *Chlamydomonas reinhardtii* as an experimental system to study chloroplast RNA metabolism. *Naturwissenschaften* 87: 97-107.
- Niefeld JJ, Van der Kraan I & 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-Harminen 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.
- 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. *J Biochem* 268: 5300-5309.
- Norman KR & Moerman DG (2000) The *let-268* locus of *Caenorhabditis elegans* encodes a procollagen lysyl hydroxylase that is essential for type IV collagen secretion. *Dev Biol* 227: 690-705.
- Novelli J, Ahmed S & Hodgkin J (2004) Gene interactions in *Caenorhabditis elegans* define DPY-31 as a candidate procollagen C-proteinase nad SQT-3/ROL-4 as its predicted major target. *Genetics* 168: 1259-1273.
- Novelli J, Page AP & Hodgkin J (2006) The C terminus of collagen SQT-3 has complex and essential functions in nematode collagen assembly. *Genetics* 172: 2253-2267.

- Nyström J, Shen ZZ, Aili M, Flemming AJ, Leroi A & Tuck S (2002) Increased or decreased levels of *Caenorhabditis elegans lon-3*, a gene encoding a collagen, cause reciprocal changes in body length. *Genetics* 161: 83-97.
- O'Kane CJ (2003) Modelling human diseases in *Drosophila* and *Caenorhabditis*. *Semin Cell Dev Biol* 14: 3-10.
- Page AP (2001) The nematode cuticle: synthesis, modification and mutants. In: Kennedy MW & Harnett W (eds) *Parasitic Nematodes*. CABI Press, United Kingdom, 167-193.
- Page AP & Winter AD (2003) Enzymes involved in the biogenesis of the nematode cuticle. *Adv Parasitol* 53: 85-148.
- Pajunen L, Jones TA, Helaakoski T, Pihlajaniemi T, Solomon E, Sheer D & Kivirikko KI (1989) Assignment of the gene coding for the α -subunit of prolyl 4-hydroxylase to human chromosome region 10q21.3-23.1. *Am J Hum Genet* 45: 829-834.
- Pan T, Groger H, Schmid V & Spring J (1998) A toxin homology domain in an astacin-like metalloproteinase of the jellyfish *Podocoryne carnea* with a dual role in digestion and development. *Dev Genes Evol* 208: 259-266.
- Pekkala M, Hieta R, Bergmann U, Kivirikko KI, Wierenga RK & Myllyharju J (2004) The peptide-substrate-binding domain of collagen prolyl 4-hydroxylases is a tetratricopeptide repeat domain with functional aromatic residues. *J Biol Chem* 279: 52255-52261.
- Pirneskoski A, Klappa P, Lobell M, Williamson RA, Byrne L, Alanen HI, Kivirikko KI, Freedman RB & Ruddock LW (2004) Molecular characterization of the principal substrate binding site of the ubiquitous folding catalyst protein disulfide isomerase. *J Biol Chem* 279: 10374-10381.
- Pirneskoski A, Ruddock LW, Klappa P, Freedman RB, Kivirikko KI & Koivunen P (2001) Domains b' and a' of protein disulfide isomerase fulfill the minimum requirement for function as a subunit of prolyl 4-hydroxylase. The N-terminal domains a and b enhance this function and can be substituted in part by those of ERp57. *J Biol Chem* 276: 11287-11293.
- 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.
- Ricard-Blum S & Ruggiero F (2005) The collagen superfamily: from the extracellular matrix to the cell membrane. *Pathol Biol* 53: 430-442.
- Riihimaa P, Nissi R, Page AP, Winter AD, Keskiäho 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, Fülöp 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.
- Roberts K, Smith MG & Hills GJ (1972) Structure, composition, and morphogenesis of the cell wall of *Chlamydomonas reinhardtii*. *Ultrastruct Res* 40: 599-613.
- Schofield CJ & Ratcliffe PJ (2005) Signalling hypoxia by HIF hydroxylases. *Biochem Biophys Res Commun* 338: 617-626.
- Schroda M (2006) RNA silencing in *Chlamydomonas*: mechanisms and tools. *Curr Genet* 49: 69-84.
- Showalter AM (1993) Structure and function of plant cell wall proteins. *Plant Cell* 5: 9-23.
- Showalter AM (2001) Arabinogalactan-proteins: structure, expression and function. *Cell Mol Life Sci* 58: 1399-1417.
- Sibley MH, Johnson JJ, Mello CC & Kramer JM (1993) Genetic identification, sequence, and alternative splicing of the *Caenorhabditis elegans* $\alpha 2(IV)$ gene. *J Cell Biol* 111: 1685-1699.

- Simmer F, Moorman C, van der Linden AM, Kuijk E, van den Berghe PV, Kamath RS, Fraser AG, Ahringer J & Plasterk RH (2003) Genome-wide RNAi of *C. elegans* using the hypersensitive *rrf-3* strain reveals novel gene functions. *PLoS Biol* 1: E12.
- Smith MC, Burns N, Sayers JR, Sorrell JA, Casjens SR & Hendrix RW (1998) Bacteriophage collagen. *Science* 279: 1834.
- Stein LD, Bao ZR, Blasiar D, Blumenthal T, Brent MR, Chen NS, Chinwalla A, Clarke L, Clee C, Coghlan A, Coulson A, D'Eustachio P, Fitch DHA, Fulton LA, Fulton RE, Griffiths-Jones S, Harris TW, Hillier LW, Kamath R, Kuwabara PE, Mardis ER, Marra MA, Miner TL, Minx P, Mullikin JC, Plumb RW, Rogers J, Schein JE, Sohrmann M, Spieth J, Stajich JE, Wei CC, Willey D, Wilson RK, Durbin R & Waterston RH (2003) The genome sequence of *Caenorhabditis briggsae*: a platform for comparative genomics. *PLOS Biol* 1: 166.
- Sulston JE & Horvitz HR (1977) Post-embryonic cell lineages of the nematode, *Caenorhabditis elegans*. *Dev Biol* 56: 110-156
- 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* 61: 188-198
- Taylor TC, Backlund A, Bjorhall K, Spreitzer RJ & Andersson I (2001) First crystal structure of Rubisco from a green alga, *Chlamydomonas reinhardtii*. *J Biol Chem* 276: 48159-48164.
- Thacker C, Peters K, Srayko M & Rose AM (1995) The *bli-4* locus of *Caenorhabditis elegans* encodes structurally distinct *kex2*/subtilisin-like endoproteases essential for early development and adult morphology. *Genes Dev* 9: 956-971.
- Thacker C, Sheps JA & Rose AM (2006) *Caenorhabditis elegans dpy-5* is a cuticle procollagen processed by a proprotein convertase. *Cell Mol Life Sci* 63: 1193-1204.
- Thein MC, McCormack G, Winter AD, Johnstone IL, Shoemaker CB & Page AP (2003) *Caenorhabditis elegans* exoskeleton collagen COL-19: an adult-specific marker for collagen modification and assembly, and the analysis of organismal morphology. *Dev Dyn* 226: 523-539.
- Tiainen P, Myllyharju J & Koivunen P (2005) Characterization of a second *Arabidopsis thaliana* prolyl 4-hydroxylase with distinct substrate specificity. *J Biol Chem* 280: 1142-1148.
- Tian G, Xiang S, Noiva R, Lennarz WJ & Schindelin H (2006) The crystal structure of yeast protein disulfide isomerase suggests cooperativity between its active sites. *Cell* 124: 61-73.
- Timmons L, Court DL & Fire A (2001) Ingestion of bacterially expressed dsRNA can produce specific and potent genetic interference in *Caenorhabditis elegans*. *Gene* 263: 103-112.
- Timmons L, Hiroaki T, Mello GC & Fire AZ (2003) Inducible systemic RNA silencing in *Caenorhabditis elegans*. *Mol Biol Cell* 14: 2972-2983.
- Valegård K, van Scheltinga AC, Lloyd MD, Hara T, Ramaswamy S, Perrakis A, Thompson A, Lee HJ, Baldwin JE, Schofield CJ, Hajdu J & Andersson I (1998) Structure of a cephalosporin synthase. *Nature* 394: 805-809.
- 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.
- van der Keyl H, Kim H, Espey R, Oke CV & Edwards MK (1994) *Caenorhabditis elegans sqt-3* mutants have mutations in the *col-1* collagen gene. *Dev Dyn* 201: 86-94.
- Vastenhouw NL, Brunschwig K, Okihara KL, Müller F, Tijsterman M & Plasterk RHA (2006) Long-term gene silencing by RNAi. *Nature* 442: 882
- 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 $\alpha\beta$ dimer with the human protein disulfide isomerase/ β subunit. *J Biol Chem* 269: 26746-26753.
- Veit G, Kobbe B, Keene DR, Paulsson M, Koch M & Wagener R (2005) Collagen XXVIII, a novel VWA-domain-containing protein with many imperfections in the collagenous domain. *J Biol Chem* 281, 3494-3504.
- von Mende N, Bird DM, Alber PS & Riddle DL (1988) *dpy-13*: a nematode collagen gene that affects body shape. *Cell* 55: 567-576.
- Vuori K, Pihlajaniemi T, Marttila M & Kivirikko KI (1992a) 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 (1992b) 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.
- Vuorela K, Myllyharju J, Nissi R, Pihlajaniemi T & Kivirikko KI (1997) Assembly of human prolyl 4-hydroxylase and type III collagen in the yeast *Pichia pastoris*: formation of a stable enzyme tetramer requires coexpression with collagen and assembly of a stable collagen requires coexpression with prolyl 4-hydroxylase. *EMBO J* 16: 6702-6712.
- Wilkinson B & Gilbert HF (2004) Protein disulfide isomerase. *Biochim Biophys Acta* 1699: 35-44.
- Williams DC, Boulin T, Ruaud AF, Jorgensen EM & Bessereau JL (2001) Characterization of Mos1-mediated mutagenesis in *Caenorhabditis elegans*: a method for the rapid identification of mutated genes. *Genetics* 169: 1779-1785.
- Wilmouth RC, Turnbull JJ, Welford RW, Clifton IJ, Prescott AG & Schofield CJ (2002) Structure and mechanism of anthocyanidin synthase from *Arabidopsis thaliana*. *Structure* 10: 93-103.
- Winter AD, Eschenlauer SC, McCormack G & Page AP (2007) Loss of secretory pathway FK506-binding proteins results in cold-sensitive lethality and associated extracellular matrix defects in the nematode *Caenorhabditis elegans*. *J Biol Chem*, in press.
- 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.
- Woessner JP & Goodenough UW (1989) Molecular characterization of a zygote wall protein: an extensin-like molecule in *Chlamydomonas reinhardtii*. *Plant Cell* 1: 901-911.
- Yuasa K, Toyooka K, Fukuda H & Matsuoka K (2005) Membrane-anchored prolyl hydroxylase with an export signal from the endoplasmic reticulum. *Plant J* 41: 81-94.
- Zhang Z, Ren J, Stammers DK, Baldwin JE, Harlos K & Schofield CJ (2000) Structural origins of the selectivity of the trifunctional oxygenase clavaminic acid synthase. *Nat Struct Biol* 7: 127-33.

Original articles

- I Keskiaho K, Hieta R, Sormunen R & Myllyharju J (2007) *Chlamydomonas reinhardtii* has multiple prolyl 4-hydroxylases, one of which is essential for proper cell wall assembly. *Plant Cell* 19: 256-269.
- II Winter AD, Keskiaho K, Kukkola L, McCormack G, Felix M-A, Myllyharju J & Page AP (2007) Differences in collagen prolyl 4-hydroxylase assembly between two *Caenorhabditis* nematode species despite high amino acid sequence identity of the enzyme subunits. *Matrix Biol*, 2007, Jan 30.
- III Keskiaho-Saukkonen K, Kukkola L, Page AP, Winter AD, Vuoristo J, Nissi R, Riihimaa P & Myllyharju J. Characterization of a novel *Caenorhabditis elegans* prolyl 4-hydroxylase with a unique substrate specificity and restricted expression in the pharynx and excretory duct. Manuscript.

The article that appeared in the *Plant Cell* (2007;19:256-269) is reproduced with permission. The copyright holder is American Society of Plant Biologists.

The article that appeared in the *Matrix Biology* (2007;Jan 30) is reproduced with permission. The copyright holder is Elsevier.

Original publications are not included in the electronic version of the dissertation.

ACTA UNIVERSITATIS OULUENSIS
SERIES D MEDICA

915. Kinnunen, Tuija (2007) Keuhkohtaumataudin sairaalahoito Suomessa: hoitoajan pituus ja sen yhteys ennusteeseen
916. Pursiainen, Ville (2007) Autonomic dysfunction in early and advanced Parkinson's disease
917. Paldanius, Mika (2007) Serological studies on *Chlamydia pneumoniae* infections
918. Parkkila, Timo (2007) Sutter metacarpophalangeal arthroplasty in rheumatoid patients
919. Hurtig, Tuula (2007) Adolescent ADHD and family environment—an epidemiological and clinical study of ADHD in the Northern Finland 1986 Birth Cohort
920. Takaluoma, Kati (2007) Lysyl hydroxylases. Studies on recombinant lysyl hydroxylases and mouse lines lacking lysyl hydroxylase 1 or lysyl hydroxylase 3
921. Majamaa-Voltti, Kirsi (2007) Cardiovascular abnormalities in adult patients with the 3243A>G mutation in mitochondrial DNA
922. Kangas, Jarmo (2007) Outcome of total Achilles tendon rupture repair, with special reference to suture materials and postoperative treatment
923. Annunen-Rasila, Johanna (2007) Molecular and cell phenotype changes in mitochondrial diseases
924. Suhonen, Marjo (2007) Osallistujaohjaus ristipaineiden keskellä. Tapaustutkimus Kainuun maakuntakokeilun sosiaali- ja terveydenhuollon kehittämishankkeen suunnitteluvaiheesta vuosina 2003–2004
925. Ylipalosaari, Pekka (2007) Infections in intensive care; epidemiology and outcome
926. Rapakko, Katrin (2007) Hereditary predisposition to breast cancer—evaluation of candidate genes
927. Peltoniemi, Outi-Maria (2007) Corticosteroid treatment in the perinatal period. Efficacy and safety of antenatal and neonatal corticosteroids in the prevention of acute and long-term morbidity and mortality in preterm infants
928. Rahko, Eeva (2007) Evaluation of tumor suppressor gene p53, oncogene c-erbB-2 and matrix-metalloproteinase-9 as prognostic and predictive factors in breast carcinoma
929. Chi, Lijun (2007) Role of Sprouty and Cerberus proteins in urogenital system development

Book orders:
OULU UNIVERSITY PRESS
P.O. Box 8200, FI-90014
University of Oulu, Finland

Distributed by
OULU UNIVERSITY LIBRARY
P.O. Box 7500, FI-90014
University of Oulu, Finland

S E R I E S E D I T O R S

A
SCIENTIAE RERUM NATURALIUM
Professor Mikko Siponen

B
HUMANIORA
Professor Harri Mantila

C
TECHNICA
Professor Juha Kostamovaara

D
MEDICA
Professor Olli Vuolteenaho

E
SCIENTIAE RERUM SOCIALIUM
Senior Assistant Timo Latomaa

E
SCRIPTA ACADEMICA
Communications Officer Elna Stjerna

G
OECONOMICA
Senior Lecturer Seppo Eriksson

EDITOR IN CHIEF
Professor Olli Vuolteenaho

EDITORIAL SECRETARY
Publications Editor Kirsti Nurkkala

ISBN 978-951-42-8472-4 (Paperback)

ISBN 978-951-42-8473-1 (PDF)

ISSN 0355-3221 (Print)

ISSN 1796-2234 (Online)

