

*Päivi Fonsén*

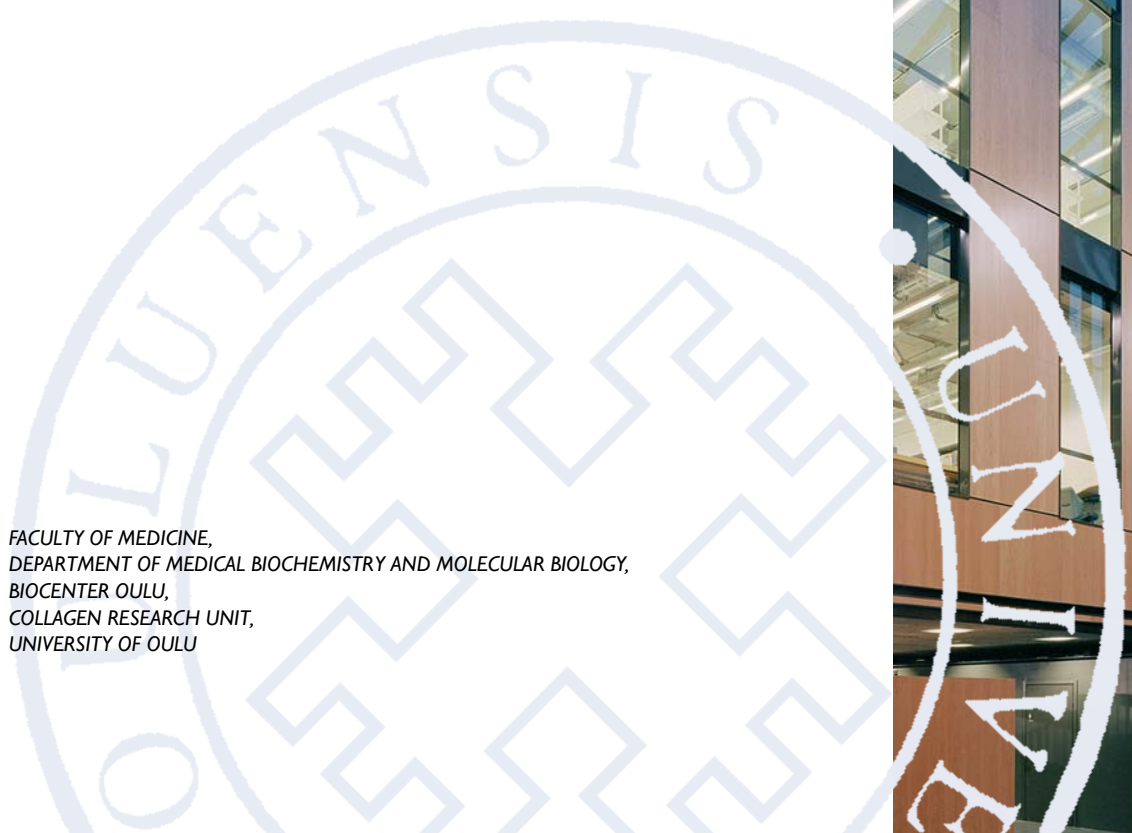
## PROLYL HYDROXYLASES

**CLONING AND CHARACTERIZATION OF NOVEL  
HUMAN AND PLANT PROLYL 4-HYDROXYLASES,  
AND THREE HUMAN PROLYL 3-HYDROXYLASES**

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

D

MEDICA





**PÄIVI FONSEN**

**PROLYL HYDROXYLASES**

Cloning and characterization of novel human and plant prolyl 4-hydroxylases, and three human prolyl 3-hydroxylases

Academic dissertation to be presented, with the assent of the Faculty of Medicine of the University of Oulu, for public defence in Auditorium F101 of the Department of Physiology (Aapistie 7), on December 21st, 2007, at 10 a.m.

Copyright © 2007  
Acta Univ. Oul. D 957, 2007

Supervised by  
Professor Johanna Myllyharju  
Docent Peppi Karppinen

Reviewed by  
Professor Joachim Fandrey  
Professor Dörthe Katschinski

ISBN 978-951-42-8672-8 (Paperback)  
ISBN 978-951-42-8673-5 (PDF)  
<http://herkules.oulu.fi/isbn9789514286735/>  
ISSN 0355-3221 (Printed)  
ISSN 1796-2234 (Online)  
<http://herkules.oulu.fi/issn03553221/>

Cover design  
Raimo Ahonen

OULU UNIVERSITY PRESS  
OULU 2007

# Fonsén, Päivi, Prolyl hydroxylases. Cloning and characterization of novel human and plant prolyl 4-hydroxylases, and three human prolyl 3-hydroxylases

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 957, 2007*  
Oulu, Finland

## *Abstract*

Prolyl hydroxylases catalyze the post-translational formation of 3- and 4-hydroxyprolines in polypeptides. To date, two prolyl 4-hydroxylase families are known to exist: collagen prolyl 4-hydroxylases (C-P4Hs) which reside in the endoplasmic reticulum, and hypoxia-inducible factor prolyl 4-hydroxylases (HIF-P4Hs) which are located in either the cytoplasm or nucleus. C-P4Hs and HIF-P4Hs belong to the 2-oxoglutarate dioxygenase family and require  $\text{Fe}^{2+}$ , 2-oxoglutarate,  $\text{O}_2$  and ascorbate in their reaction. C-P4Hs are critical enzymes in collagen biosynthesis since the formation of 4-hydroxyproline residues stabilizes the collagen triple helix at body temperature. HIF-P4Hs regulate, through hypoxia-inducible factor HIF, the expression of genes that are essential for the survival of cells during hypoxia.

This thesis reports the cloning and characterization of two novel P4Hs, from human and a plant, which show some distinct features when compared to previously characterized P4Hs. The human P4H was found to have a unique transmembrane domain, with its catalytic region residing inside the lumen of the endoplasmic reticulum, its subcellular localization thus being identical to that of the C-P4Hs. However, unlike C-P4Hs, it hydroxylated HIF- $\alpha$  both *in vitro* and *in cellulo*. Furthermore, its expression level was induced in hypoxic conditions in most of the cell lines studied. The *Arabidopsis thaliana* P4H had distinct differences in its substrate specificity when compared to another previously characterized *A. thaliana* P4H. Interestingly, the putative peptide binding regions of the two new P4Hs characterized in this study shared some homology.

Three prolyl 3-hydroxylase (P3H) isoenzymes are known to exist in vertebrates and they also belong to the 2-oxoglutarate dioxygenases. It is known that 3-hydroxyprolines are found only in collagens, being most abundant in type IV collagen. The function of this modification is as yet poorly understood, but its absence in collagen I has recently been shown to lead to recessive lethal osteogenesis imperfecta. The human P3H isoenzymes were cloned during these thesis studies, and were expressed as recombinant proteins. The kinetic properties of one of them, P3H2, which was found to be expressed in structures rich with basement membranes, was studied in detail.

*Keywords:* collagen, hypoxia-inducible factor, prolyl 3-hydroxylase, prolyl 4-hydroxylase



# Fonsén, Päivi, Prolyylihydroksylaasit kollageenisynteesissä ja elimistön vasteissa vähähappisiin olosuhteisiin

Lääketieteellinen tiedekunta, Lääketieteellisen biokemian ja molekyylibiologian laitos, Biocenter Oulu, Kollageenitutkimusyksikkö, Oulun yliopisto, PL 5000, 90014 Oulun yliopisto  
*Acta Univ. Oul. D 957, 2007*  
Oulu

## Tiivistelmä

Prolyylihydroksylaasit ovat entsyymejä, jotka katalysoivat 3- ja 4-hydroksiproliniin muodostumisen valkuaisaineissa. Nykyisin tunnetaan ainakin kaksi prolyyli-4-hydroksylaasin (P4H) entsyymiperhettä: endoplasmakalvostossa sijaitsevat kollageeni prolyyli-4-hydroksylaasit (kollageeni-P4H:t) sekä vähähappisissa olosuhteissa aktivoituvaa transkriptiotekijää, hypoksiainduoituvaa faktoria (HIF), hydroksyloivat prolyyli-4-hydroksylaasit (HIF-P4H:t). HIF-P4H:t sijaitsevat sytoplasmassa ja tumassa. Sekä kollageeni-P4H:t että HIF-P4H:t kuuluvat 2-oksoglutaraattidioksygenaasin laajaan entsyymiperheeseen. Nämä entsyymit tarvitsevat kosubstraateikseen rautaa, 2-oksoglutaraattia, happea sekä C-vitamiinia. Kollageeni-P4H:t hydroksyloivat kollageenin proliinitähteitä ja ovat avainasemassa kollageenisynteesissä, sillä muodostuneet 4-hydroksiprolinitähteet ovat ehdoton vaatimus stabiilille kollageenirakenteelle. HIF-P4H:t säätelevät puolestaan niiden geenien ilmenemistä, jotka ovat välttämättömiä elimistön selviytymiselle vähähappisissa olosuhteissa. HIF-P4H:t hydroksyloivat HIF-transkriptiotekijän  $\alpha$ -alayksikön tiettyjä proliinitähteitä hapen läsnä ollessa, joka ohjaa  $\alpha$ -alayksikön proteasomaaliseen hajotukseen eikä aktiivista HIF transkriptiotekijää siten muodostu. Alentuneessa happipitoisuudessa HIF-P4H entsyymien toiminta estyy, HIF stabiloituu ja aktivoi kohdegeeniensä toiminnan. Kollageeni-P4H entsyymejä pidetään erityisen sopivina lääkekehityksen kohteina fibroottisten ja HIF-P4H entsyymejä iskeemisten sairauksien hoitoon.

Tässä väitöskirjatyössä on karakterisoitu aiemmin tuntematon ihmisen transmembraaninen P4H entsyymi (P4H-TM). Entsyymi osoittautui indusoituvan vähähappisissa olosuhteissa useissa solulinjoissa ja hydroksyloivan HIF-transkriptiotekijää muistuttaen siten HIF-P4H entsyymejä. Kuitenkin P4H-TM:n solulokalisatio poikkesi HIF-P4H entsyymeistä, sillä sen havaittiin sijaitsevan endoplasmakalvostossa, katalyyttinen keskus kalvoston sisällä. Näiden tutkimustulosten valossa on oletettavaa, että tällä ihmisentsyymillä on HIF:n lisäksi toinen toistaiseksi tuntematon fysiologinen substraatti. Väitöskirjassa karakterisoiitiin toinen lituruohon, Arabidopsis thalianan, P4H (At-P4H-2), joka poikkesi katalyyttisiltä ominaisuuksiltaan aiemmin karakterisoidusta lituruohon P4H:sta. Näiden kahden kasvientsyymien substraattivaatimusten poiketessa selvästi toisistaan, on niillä solussa todennäköisesti spesifiset tehtävät. At-P4H-2:n oletetulla substraatin sitomisalueella on jakso, joka on 37-prosenttisesti identtinen ihmisen P4H-TM:n kanssa, minkä vuoksi At-P4H-2:n karakterisoinnin uskottiin olevan tärkeä apuväline ihmisen P4H-TM:n tutkimuksissa.

Selkärankaisilla prolyyli-3-hydroksylaaseja (P3H) tiedetään olevan kolme, ja myös ne kuuluvat 2-oksoglutaraattidioksygenaaseihin. 3-Hydroksiproliniinia esiintyy ainoastaan kollageeneissa, erityisesti tyypin IV kollageenissa, joka on tyvikalvojen tärkeä rakennekomponentti. 3-Hydroksiproliniin merkitystä ei tunneta vielä tarkoin, mutta tyypin I kollageenissa 3-hydroksiproliniin puutoksen on osoitettu johtavan vaikeaan luustosairauteen, osteogenesis imperfectaan. Väitöskirjatyössä ihmisen P3H:t kloonattiin ja tuotettiin rekombinanttiproteiineina. Yhden isoentsyymien (P3H2) katalyyttiset ominaisuudet määritettiin ja sen osoitettiin ilmenevän erityisesti kudoksissa, joissa on paljon tyvikalvorakenteita.

*Asiasanat:* hypoksiainduoituva faktori, kollageeni, prolyyli-3-hydroksylaasi, prolyyli-4-hydroksylaasi



## 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 warmest gratitude to my supervisor, Professor Johanna Myllyharju, for the opportunity to prepare my thesis in her group under her excellent supervision. Her wide knowledge of science and enthusiasm towards it has provided an excellent basis for this work. My other supervisor, Docent Peppi Karppinen, has my thanks for introducing me to the field of science from the very beginning, and for supervising me in such an excellent way. I have been fortunate to have her as my supervisor. Her never-ending optimism and enthusiasm towards my projects and science in general has carried along this work. My sincere thanks go to Academy Professor Emeritus Kari Kivirikko for his endless ideas and optimism about my projects. Professor and head of the department Taina Pihlajaniemi is acknowledged for providing excellent research facilities and an inspiring scientific atmosphere in the department and at Biocenter Oulu. I wish to express my respect for Professor Emeritus Ilmo Hassinen, Professor Leena Ala-Kokko and Professor Seppo Vainio for their outstanding scientific work.

I am very grateful to Professors Joachim Fandrey and Dörthe Katschinski for their careful review of this thesis and valuable comments on it. Dr. Peter Cura is acknowledged for his revision of the language.

I wish to thank Dr. Stephen Klaus, Docent Raija Sormunen and Dr. Kim Williams for their valuable efforts on these projects. Jaana Hyvärinen and Annika Pasanen are thanked for sharing the projects with me.

This thesis could not have been completed without the excellent and skilful technical assistance of Tanja Aatsinki, Anne Kokko, Outi Mänty, Merja Nissilä and Riitta Polojärvi. I would like to thank the whole staff of the department for their help in many scientific and practical matters especially Auli Kinnunen, Seppo Lähdesmäki, Pertti Vuokila and Marja-Leena Karjalainen. Special thanks are due to Risto Helminen for his kind and speedy help in computer-related matters.

I would like to thank all members of our group for being such pleasant friends at work. Because of you it has been nice to come to work every day. I wish to express my warmest thanks to my office mates Antje, Katriina, Maija and Minna. These years have been full of cheerful moments in the office and outside.

I also wish to thank Juha Jäälinoja for sharing the time preparing the theses and organizing the dissertation party.

My warmest thanks go to my respected parents Eija and Pertti for providing me a loving home and supporting me in so many ways during these years away from 'home'. My sister Heli and her fiancé Harri are warmly thanked for their friendship and for many hilarious moments. My mother-in-law Tuula and her lifetime-companion Seppo are thanked for their support and friendship during these years. My aunt Anneli and her husband Kauko are thanked for being my Oulu family during these years. Kaisa, Teemu and Eeva will certainly deserve big hugs for being our exercise and excursion mates here in Oulu. Special thanks belong to the wonderful women in 'opintopiiri' which later took the role of 'h-piiri'- Anna K, Anna H, Annemari, Saana and Kaisa. You are true friends and have cheered up every single day during my medical studies and this thesis work. Without you, I doubt that it would have been possible to carry out this thesis work and medical studies simultaneously. H-piiri rules forever! All my other friends and their families in Oulu and down South deserve my big thanks for being there whenever needed. Without friends, the world would be a boring and grey place to live!

Above all, I owe my deepest and warmest gratitude to my beloved husband Mika and our wonderful daughter Lotta. Mika is thanked for being part of my and Lotta's everyday life. His love for Lotta and me and his encouragement has carried me along during these years. And not to forget your unbelievable computer skills! Lotta deserves hugs and kisses for being the most precious and happiest creature in the world. You both make me love you more and more every day.

This work was supported by the Finnish Centre of Excellence Programme 2000-2005 of the Academy of Finland, the Biocenter Oulu, the Sigrid Juselius Foundation and the Farnos Science and Research Foundation.

Oulu, September 2007

Päivi Fonsén

## Abbreviations

3Hyp	3-hydroxyproline
4Hyp	4-hydroxyproline
AGP	arabinogalactan protein
At-P4H	<i>Arabidopsis thaliana</i> prolyl 4-hydroxylase
bp(s)	base pairs
cDNA	complementary DNA
C-P4H	collagen prolyl 4-hydroxylase
Cr-P4H-1	<i>Chlamydomonas reinhardtii</i> prolyl 4-hydroxylase 1
CRTAP	cartilage associated protein
CTAD	C-terminal transactivation domain
ER	endoplasmic reticulum
FIH	factor inhibiting HIF
HIF	hypoxia-inducible factor
HIF-P4H	HIF prolyl 4-hydroxylase
HRE	hypoxia response element
HRGP	hydroxyproline-rich glycoprotein
K <sub>i</sub>	inhibitory constant
K <sub>m</sub>	Michaelis-Menten constant
mRNA	messenger RNA
LH	lysyl hydroxylase
NTAD	N-terminal transactivation domain
ODDD	oxygen-dependent degradation domain
OI	osteogenesis imperfecta
P3H	prolyl 3-hydroxylase
P4H	prolyl 4-hydroxylase
P4H-TM	prolyl 4-hydroxylase with a transmembrane domain
PBCV-1	<i>Paramecium bursaria</i> <i>Chlorella</i> virus-1
PCR	polymerase chain reaction
PDI	protein disulfide isomerase
PHY	<i>Caenorhabditis elegans</i> collagen prolyl 4-hydroxylase $\alpha$ subunit
PRP	proline-rich protein
RACE	rapid amplification of cDNA ends
siRNA	small interfering RNA
TPR	tetratricopeptide repeat
VHL	von Hippel-Lindau

X, in -Gly-X-Y- any amino acid

Y, in -Gly-X-Y- any amino acid

## List of original articles

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

- I Koivunen P, Tiainen P\*, Hyvärinen J, Williams KE, Sormunen R, Klaus SJ, Kivirikko KI & Myllyharju J (2007) An endoplasmic reticulum transmembrane prolyl 4-hydroxylase is induced by hypoxia and acts on hypoxia-inducible factor  $\alpha$ . *J Biol Chem* 282: 30544-30552.
- II 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.
- III Tiainen P\*, Pasanen A, Sormunen R & Myllyharju J (2007) Characterization of recombinant human prolyl 3-hydroxylase isoenzyme 2, a basement membrane enzyme. Submitted.

\*Fonsén née Tiainen



# Contents

<b>Abstract</b>	
<b>Tiivistelmä</b>	
<b>Acknowledgements</b>	<b>7</b>
<b>Abbreviations</b>	<b>9</b>
<b>List of original articles</b>	<b>11</b>
<b>Contents</b>	<b>13</b>
<b>1 Introduction</b>	<b>15</b>
<b>2 Review of the literature</b>	<b>17</b>
2.1 Occurrence of 4-hydroxyproline in animal and plant proteins.....	17
2.1.1 Collagens and related proteins.....	17
2.1.2 Hypoxia-inducible factor (HIF).....	19
2.1.3 Plant proteins.....	22
2.2 Collagen prolyl 4-hydroxylases .....	24
2.2.1 Vertebrate collagen prolyl 4-hydroxylases .....	25
2.2.2 Invertebrate collagen prolyl 4-hydroxylases .....	32
2.3 Plant and viral prolyl 4-hydroxylases .....	34
2.3.1 Higher plant and algal prolyl 4-hydroxylases.....	34
2.3.2 <i>Paramecium bursaria</i> <i>Chlorella</i> virus-1 prolyl 4- hydroxylase .....	35
2.4 HIF prolyl 4-hydroxylases .....	36
2.5 Prolyl 3-hydroxylases .....	39
<b>3 Outlines of the present research</b>	<b>43</b>
<b>4 Materials and methods</b>	<b>45</b>
<b>5 Results</b>	<b>47</b>
5.1 Cloning and characterization of a novel ER transmembrane human prolyl 4-hydroxylase, P4H-TM (I).....	47
5.1.1 Cloning and expression of recombinant P4H-TM.....	47
5.1.2 P4H-TM is expressed in many human tissues and cell lines and is induced by hypoxia.....	48
5.1.3 Actions of P4H-TM closely resemble those of HIF-P4Hs .....	49
5.2 Molecular cloning and characterization of the second <i>Arabidopsis thaliana</i> prolyl 4-hydroxylase, At-P4H-2 (II).....	51
5.2.1 Cloning of At-P4H-2 .....	51
5.2.2 Expression of recombinant At-P4H-2 in insect cells.....	52

5.2.3	Substrate specificity and catalytic properties of recombinant At-P4H-2.....	52
5.3	Characterization of the human prolyl 3-hydroxylase isoenzyme 2 (III).....	54
5.3.1	Cloning and expression of the recombinant P3H isoenzymes in insect cells.....	54
5.3.2	Substrate specificity and catalytic properties of recombinant P3H2.....	55
5.3.3	Expression of the P3H isoenzymes in vertebrate tissues.....	56
<b>6</b>	<b>Discussion</b>	<b>59</b>
6.1	The actions of P4H-TM resemble those of HIF-P4Hs but it may also act on substrates other than HIF-1 $\alpha$ .....	59
6.2	At-P4H-2 and At-P4H1 have distinct substrate specificities.....	61
6.3	P3H2 is associated with the hydroxylation of basement membrane collagens.....	62
<b>7</b>	<b>Conclusions and future prospects</b>	<b>65</b>
	<b>References</b>	<b>67</b>
	<b>Original articles</b>	<b>81</b>

# 1 Introduction

Collagens are the most abundant proteins in the extracellular matrix and have many important structural and functional roles in tissues. The stability of collagens is dependent on the presence of 4-hydroxyproline residues. 4-Hydroxyprolines also play a vital role in the regulation of oxygen homeostasis by targeting hypoxia-inducible factor (HIF) to proteasomal degradation in normoxic conditions. The generation of 4-hydroxyproline residues is catalyzed by two different prolyl 4-hydroxylase (P4H) families: collagen P4Hs (C-P4Hs) which reside in the endoplasmic reticulum (ER), and HIF-P4Hs which reside in the cytoplasm or nucleus. Both C-P4Hs and HIF-P4Hs belong to the 2-oxoglutarate dioxygenase family and require  $\text{Fe}^{2+}$ , 2-oxoglutarate,  $\text{O}_2$  and ascorbate. In this study a novel human P4H with a transmembrane domain (P4H-TM) was cloned and expressed as a recombinant protein. Characterization of P4H-TM revealed that it shared some features with both C-P4Hs and HIF-P4Hs as its catalytic sites were located within the lumen of the ER, however, it hydroxylated HIF instead of collagen polypeptides. The data obtained suggested that it may also have some other physiological substrates, however. A second *Arabidopsis thaliana* P4H (At-P4H-2) was cloned and characterized during these thesis studies, as it shared a marked homology with the human P4H-TM in the putative peptide substrate binding site and could thus be a potentially useful tool for further analysis of the substrate binding properties of P4H-TM. At-P4H-2 was found to have substrate specificity distinct from that of the previously characterized At-P4H-1.

Another enzyme group that also belongs to the 2-oxoglutarate dioxygenase family is the prolyl 3-hydroxylase (P3Hs) family, which has three members in vertebrates. P3Hs catalyze the formation of 3-hydroxyproline in collagens, being most abundantly found in type IV collagen. Only very little is known about the function of 3-hydroxyproline residues. The most recent data suggest that they are involved in the correct assembly of collagen molecules and collagen-protein interactions, and that they have an important role since a lack of them leads to osteogenesis imperfecta. One of the goals of this thesis work was to clone the three human P3H isoenzymes and express them as recombinant proteins so that the catalytic properties of this poorly characterized enzyme family could be studied.



## 2 Review of the literature

### 2.1 Occurrence of 4-hydroxyproline in animal and plant proteins

4-Hydroxyprolines are found in certain animal and plant proteins where they have important functions in either maintaining the structural integrity of the protein or contributing to cell survival in hypoxic conditions. The majority of 4-hydroxyproline in animal proteins is found in collagens and in other proteins containing collagen-like sequences. The stability of HIF is regulated through 4-hydroxyproline residues. 4-Hydroxyproline is also found in plants in hydroxyproline-rich glycoproteins (HRGPs) that provide structural integrity for cell walls and have several other important roles in plant cell growth and development.

#### 2.1.1 Collagens and related proteins

Collagens are the most abundant proteins in the human body and they are found in all tissues. Their main function is to maintain the structural integrity of tissues but they are also involved in a number of other biological functions including cell adhesion, chemotaxis, migration, the dynamic interplay between cells, and tissue remodelling. Currently 28 different collagen types composed of 43 distinct collagen polypeptide chains are known. In addition, there are more than 20 proteins that contain collagenous triple-helical domains. (Kielty & Grant 2002, Myllyharju & Kivirikko 2001, 2004; Ricard-Blum & Ruggiero 2005, Veit *et al.* 2006)

The diverse family of collagens can be divided into eight subfamilies depending on their sequence, structure and supramolecular assembly: 1) fibril forming collagens (types I, II, III, V, XI, XXIV, XXVII); 2) fibril associated collagens with interrupted triple helices and related collagens (IX, XII, XIV, XVI, XIX, XX, XXI, XXII and XXVI); 3) collagens that form hexagonal networks (VIII and X); 4) the family of type IV collagens of basement membranes; 5) type VI collagen, which forms beaded filaments; 6) type VII collagen, which forms anchoring fibrils; 7) collagens with transmembrane domains (types XIII, XVII, XXIII, XXV) and 8) the multiplexin subfamily (types XV and XVIII). The sequence of type XXVIII collagen is unique and it cannot be classified into any of these subfamilies even though it has similarities to collagen VI and its localization

in basement membranes could also indicate a relationship with collagen IV. (Myllyharju & Kivirikko 2004, Veit *et al.* 2006)

All collagen molecules consist of three polypeptide chains (called  $\alpha$  chains) that are each coiled into a left-handed helix and the three chains are wrapped around each other to form a right-handed superhelical structure. Collagen chains can associate into homotrimers or heterotrimers formed from identical or differing  $\alpha$  chains, respectively. Collagenous domains consist of distinct -Gly-X-Y- repeats and at least one collagenous domain is found in all collagens, the number ranging from a few dozen to 510. Every third amino acid is a glycine – this is small enough to fit in the restricted space within the helix thus permitting the triple-helical packing. Proline is typically found in the X position and 4-hydroxyproline in the Y position. In addition to collagenous domains, each collagen also has non-collagenous domains, which have several important functions. (for reviews, see Myllyharju & Kivirikko 2004, Koide & Nagata 2005, Myllyharju 2005, Ricard-Blum & Ruggiero 2005)

Collagen chains go through several post-translational modifications, which include prolyl 4-hydroxylation, prolyl 3-hydroxylation, lysyl hydroxylation and glycosylation, before assembly of the triple helix. The presence of 4-hydroxyproline residues in the Y positions is required for the thermal stability of collagen. Non-hydroxylated type I collagen is denatured at 24°C while the hydroxylated molecules are stable at temperatures up to 39°C (Berg & Prockop 1973, Jimenez *et al.* 1973). 4-Hydroxyproline is found in vertebrate collagens only in the Y positions of the -Gly-X-Y- repeats, and its presence in the X position causes destabilization of collagen (Inouye *et al.* 1982). Both proline and 4-hydroxyproline are important for the correct conformation and stability of collagen since their rigid ring structure prevents rotation around the N-C peptide bond (for a review, see Engel & Bächinger 2005). The previously proposed theory that water-bridges formed between the OH-group of 4-hydroxyproline and backbone groups provide a stabilizing effect (Bella *et al.* 1994, 1995) is no longer valid since the stabilization is also observed under anhydrous conditions and 4-hydroxyproline can be replaced with an electronegative 4(R)-fluoroproline (for a review, see Engel & Bächinger 2005). It has therefore been proposed that the 4-hydroxyproline residues stabilize the pyrrolidine ring's puckering and preorganize all three main chain torsion angles (Jenkins & Raines 2002). The triple helix is also stabilized by interchain hydrogen bonds between the backbone NH-group of a glycine and the backbone CO of a residue in the X position of a neighboring chain (for a review, see Engel & Bächinger 2005). The number of 4-

hydroxyproline residues varies within a narrow range between different collagen types and the number of residues correlates with the thermal stability (Burjanadze 1979). The best studied type I collagen contains 100 4-hydroxyproline residues per 1000 amino acids (for a review, see Kivirikko & Pihlajaniemi 1998).

In addition to collagens, collagen-like sequences are present in more than 20 vertebrate proteins that are not classified as collagens. This is because, at least at the time of their discovery, they were not known to have any role in the extracellular matrix. This heterogeneous group includes, for example, a subcomponent C1q of complement, a C1q-like factor, the hormone adiponectin, the tail structure of acetylcholinesterase, three macrophage receptors, ectodysplasin, two elastic fibre-associated glycoproteins, a src-homologous-and-collagen protein, and at least eight collectins and three ficolins (humoral lectins of the innate immune defence system). A single 4-hydroxyproline residue is present in the hydroxyproline-lysyl-bradykinin and hydroxyproline luteinizing hormone-releasing hormone. (for reviews, see Kivirikko *et al.* 1992, Kivirikko & Pihlajaniemi 1998, Myllyharju & Kivirikko 2004, Myllyharju 2005, Ricard-Blum & Ruggiero 2005)

4-Hydroxyproline is also present in elastin, a rubber-like protein that is the main component of the elastic fibers that provide the elastic properties of the extracellular matrix (for a review, see Rodgers & Weiss 2005). Elastin contains repeating -Gly-X-Y- sequences, but it has no collagen-like triple-helical domains. Its 4-hydroxyproline content varies greatly from about 10 to 50 residues per 1000 amino acids (for reviews, see Kivirikko & Pihlajaniemi 1998, Myllyharju 2005).

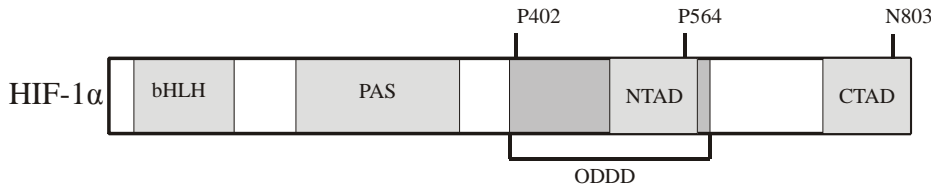
Several collagen types are also found in invertebrates. The nematode *Caenorhabditis elegans*, for example, has about 175 genes which code for cuticle collagen polypeptide chains and three genes for basement membrane collagens (for reviews, see Johnstone 2000, Page 2001, Page & Winter 2003, Myllyharju & Kivirikko 2004, Myllyharju 2005). In contrast to the large number of collagens found in *C. elegans*, *Drosophila melanogaster* has only three collagen genes, all of them encoding basement membrane collagens, and one gene for pericardin, a protein with a collagenous domain similar to collagen IV (for reviews, see Myllyharju & Kivirikko 2004, Myllyharju 2005).

### **2.1.2 Hypoxia-inducible factor (HIF)**

An adequate oxygen level is essential for all living organisms and the cellular oxygen concentration needs to be precisely controlled by the coordinated

regulation of a variety of genes. Oxygen plays an extremely important role in physiological processes and diminished oxygen concentration in tissues is associated with several diseases, such as myocardial and cerebral infarction, atherosclerosis, anaemia, diabetes and cancer, as well as failed embryonic development. HIF is the main regulator of cellular responses to hypoxia in animals as it regulates the transcription of a multitude of genes involved in the adaptation to hypoxic conditions, such as those for angiogenesis, erythropoiesis, glucose metabolism and apoptosis. It is believed that 1-1.5% of the genome is transcriptionally regulated by hypoxia and more than 100 genes are known to be regulated by HIF. (for reviews, see Ke & Costa 2006, Lahiri *et al.* 2006, Brahimi-Horn & Pouyssegur 2007, Bruegge *et al.* 2007, Siddiq *et al.* 2007)

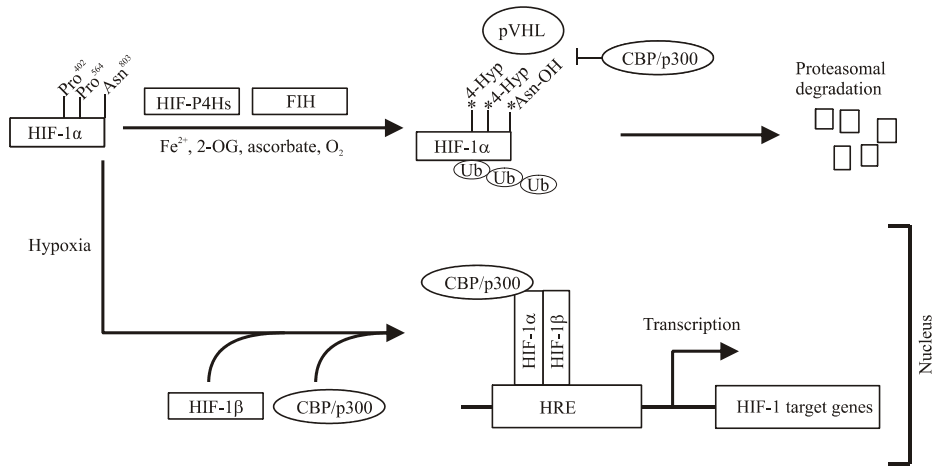
HIF is an  $\alpha\beta$  heterodimer consisting of an oxygen-regulated  $\alpha$  subunit and a constitutively expressed  $\beta$  subunit (Wang *et al.* 1995). Three HIF- $\alpha$  isoforms (HIF-1 $\alpha$  to HIF-3 $\alpha$ ) and three HIF- $\beta$  isoforms (HIF-1 $\beta$  to HIF-3 $\beta$ ) exist in humans (for reviews, see Kaelin 2005a, Ke & Costa 2006). These proteins belong to the family of basic helix-loop-helix Per-Arnt-Sim proteins (Wang *et al.* 1995), and share a number of structural and biochemical similarities. HIF-1 $\alpha$  and HIF-2 $\alpha$  are closely related to each other and contain separate domains dedicated to DNA binding and dimerization with HIF- $\beta$ , as well as two transactivation domains (NTAD and CTAD) which can activate the transcription of hypoxia-responsive genes (Fig. 1) (for reviews, see Kaelin 2005a, b; Bruegge *et al.* 2007). The little studied HIF-3 $\alpha$  exists as no less than six splice variants in humans and differs in its structure when compared to HIF-1 $\alpha$  and HIF-2 $\alpha$  (for a review, see Brahimi-Horn & Pouyssegur 2007). One of the splice variants of HIF-3 $\alpha$ , termed IPAS, is considered to be a HIF-1 $\alpha$  antagonist. This binds to HIF-1 $\alpha$  thus preventing its dimerization with HIF-1 $\beta$  into an active transcription factor (for a review, see Bruegge *et al.* 2007). Both HIF-1 $\alpha$  and HIF-2 $\alpha$  have two conserved sites for prolyl 4-hydroxylation (Pro<sup>402</sup> and Pro<sup>564</sup> in human HIF-1 $\alpha$ ) within an oxygen-dependent degradation domain (ODDD) (Fig. 1) (Srinivas *et al.* 1999, Masson *et al.* 2001, Masson & Ratcliffe 2003). In HIF-3 $\alpha$  this domain is partially conserved and contains a single site for prolyl 4-hydroxylation (for a review, see Schofield & Ratcliffe 2005).



**Fig. 1.** The domain structure of human HIF-1 $\alpha$ . HIF-1 $\alpha$  contains an oxygen-dependent degradation domain (ODDD) where two conserved proline (P) residues hydroxylated by HIF-P4Hs reside. The conserved asparagine (N) residue hydroxylated by FIH resides in the C-terminal transactivation domain (CTAD). The ODDD, the domains dedicated to dimerization and DNA binding (bHLH and PAS), and the two transactivation domains (NTAD and CTAD) are conserved in HIF-2 $\alpha$ . While all HIF-3 $\alpha$  forms lack the CTAD, some of its short forms also lack the ODDD that includes the NTAD.

The stability and transcriptional activity of HIF is regulated by two post-translational oxygen-dependent events. Under normoxia hydroxylation of one or two of the conserved proline residues in the ODDD of HIF-1 $\alpha$  generates a binding site for the von Hippel-Lindau (VHL) ubiquitin-protein ligase complex, which targets HIF-1 $\alpha$  for subsequent polyubiquitination and proteasomal degradation (Epstein *et al.* 2001, Ivan *et al.* 2001, Jaakkola *et al.* 2001, Masson *et al.* 2001, Yu *et al.* 2001). The proline residues become 4-hydroxylated in normoxia by the HIF-P4Hs (Bruick & McKnight 2001, Epstein *et al.* 2001, Ivan *et al.* 2002). There is also an additional mechanism for controlling the level of transactivation by HIF- $\alpha$  in normoxia. Hydroxylation of a conserved asparagine (Asn<sup>803</sup> in HIF-1 $\alpha$ , see Fig. 1) in the CTAD by an asparaginyl hydroxylase, known as factor inhibiting HIF (FIH) (Mahon *et al.* 2001, Hewitson *et al.* 2002, Lando *et al.* 2002a), prevents the binding of the transcriptional coactivator CBP/p300 to HIF- $\alpha$  (Lando *et al.* 2002b).

In hypoxia, the oxygen level is not sufficient for HIF-P4Hs and FIH to function. As a result, HIF- $\alpha$  escapes ubiquitination and proteasomal destruction, leading to an increased HIF- $\alpha$  protein level. Once HIF- $\alpha$  is stabilized, it can dimerize with HIF- $\beta$ , bind to the hypoxia response elements (HRE) of its target genes and induce their expression. A schematic presentation of the oxygen-dependent regulation of HIF-1 is shown in Fig. 2. (for reviews, see Fedele *et al.* 2002, Kaelin 2005a, b; Schofield & Ratcliffe 2005, Ke & Costa 2006, Semenza 2006, Siddig *et al.* 2007)



**Fig. 2. Oxygen-dependent regulation of HIF-1 stabilization and transactivation.** In normoxia the proline residues Pro<sup>402</sup> and Pro<sup>564</sup> and Asn<sup>803</sup> in HIF-1 $\alpha$  are hydroxylated by HIF-P4Hs and FIH, respectively. In addition to oxygen, the hydroxylation reaction requires Fe<sup>2+</sup>, 2-oxoglutarate (2-OG) and ascorbate. The hydroxylated proline residue(s) enable binding of the pVHL ubiquitin-ligase complex to HIF-1 $\alpha$ , targeting the HIF-1 $\alpha$  to proteasomal degradation. Hydroxylation of the Asn<sup>803</sup> blocks the recruitment of the transcriptional coactivator CBP/p300. In hypoxia, the activities of the HIF-P4Hs and FIH are inhibited due to the lack of oxygen. This leads to the stabilization and accumulation of HIF-1 $\alpha$ . The stabilized HIF-1 $\alpha$  proteins translocate into the nucleus and form active dimers with HIF-1 $\beta$ . Furthermore, the nonhydroxylated Asn<sup>803</sup> allows binding of the CBP/p300 to HIF. The complex binds to the HREs of a large number of hypoxia-inducible genes and induces their expression.

### 2.1.3 Plant proteins

Cellulose and other polysaccharides, as well as hydroxyproline-rich glycoproteins (HRGPs), are constituents of plant cell walls. As structural components, HRGPs are as important to plant cell walls as collagens are to the extracellular matrix of vertebrates. HRGPs play a wide range of functions from providing structural integrity to mediating cell-cell interactions to contributing to plant defence. The role of HRGPs in cell wall assembly is dependent on extensive post-translational modifications: prolyl 4-hydroxylation, glycosylation and crosslinking. (for reviews, see Showalter 1993, 2001; Kieliszewski & Lamport 1994, Cassab 1998)

HRGPs are usually covalently crosslinked into large meshworks, providing tensile strength to the cell walls. HRGPs account for as much as 10-20% of the

dry weight of the cell wall. They consist of multiple short repetitive sequences which are extensively glycosylated, mainly through 4-hydroxyproline residues. In higher plants the HRGP superfamily contains three major groups of proteins: the arabinogalactan proteins, extensins and proline-rich glycoproteins. (for reviews, see Cassab 1998, Kieliszewski & Shpak 2001, Showalter 2001)

Arabinogalactan proteins (AGPs) are the major components of all higher plant tissues and of many lower plant tissues as well. AGPs are structurally complex and consist of a core protein, which is typically rich in 4-hydroxyproline, alanine, serine, threonine and glycine residues, and contains -Ala-4Hyp- repeats, and carbohydrate side chains rich in arabinose and galactose polysaccharides. Carbohydrates can form up to 95% of the mass of AGPs and these are mainly O-linked to 4-hydroxyproline, but also to serine and threonine residues. It is assumed that AGPs have important roles in plant growth, development and wound healing. Soluble and highly glycosylated AGPs are located on cell membranes, but are not covalently linked to the cell wall, which implies that they do not have a structural function. (for reviews, see Kieliszewski & Shpak 2001, Cassab 1998, Showalter 2001)

Extensins are a family of plant cell wall glycoproteins that are rich in 4-hydroxyproline and serine, and some combination of the amino acids valine, tyrosine, lysine and histidine. They also contain characteristic -Ser-(4Hyp)<sub>4</sub>- sequences. Extensins have other repetitive sequences as well, and these vary from plant to plant, but have similarities with each other. Most of the 4-hydroxyproline residues are O-glycosylated with one to four arabinosyl residues, and many of the serine residues are galactosylated. Insoluble extensins are synthesized as soluble monomers that become insoluble after secretion into the cell wall, probably as a result of crosslinking. Extensins generally adopt a poly(L-proline) type II helical structure in solution. Extensins are the only HRGPs for which a specific function in the cell wall architecture is defined: they participate in crosslink scaffolds analogous to collagens. In addition to their structural role, they are involved in development, plant defence and wound healing. (for reviews, see Showalter 1993, Cassab 1998, Kieliszewski & Shpak 2001)

Lectins are carbohydrate binding proteins which exist in plant cell walls. They play a major role in cell-cell recognition and signalling, and are involved in plant defence systems. 4-Hydroxyproline-rich lectins are extensin-like proteins and are an example of chimeric extensins. They have only been isolated from solanaceous plants. The potato tuber lectin is the most extensively studied and consists of at least two domains: one rich in serine and 4-hydroxyproline that also

contains a carbohydrate moiety, and the other rich in glycine and cysteine. (for reviews, see Showalter 1993, Cassab 1998, Van Damme *et al.* 2004)

Proline-rich proteins (PRPs) are also plant cell wall proteins that form covalently crosslinked networks. They are highly periodic, with at least one characteristic -Pro-4Hyp-Val-Tyr-Lys- repeat, and lightly glycosylated proteins. They play a role in several aspects of plant development, ranging from germination to pod formation and nodule morphogenesis. It seems that PRPs become insoluble in plant cell walls in a manner similar to extensins by forming crosslinks. (for reviews, see Cassab 1998, Kieliszewski & Shpak 2001)

HRGPs are also found in the cell walls of the volvocine algae, *Chlamydomonas reinhardtii* and *Volvox carteri*. Their multilayered cell walls consist mainly of an insoluble HRGP framework and lack carbohydrate polymers. The algal HRGPs have distinct similarities to the extensins of higher plants and are therefore thought to share similar functions. The *C. reinhardtii* cell wall contains 25-30 different HRGPs. However, the outer cell wall layer contains only three of them, GP1, 2 and 3. (for reviews, see Adair & Snell 1990, Sumper & Hallmann 1998)

## **2.2 Collagen prolyl 4-hydroxylases**

Collagen prolyl 4-hydroxylases (C-P4Hs, EC 1.14.11.2) belong to the 2-oxoglutarate-dependent dioxygenase family and require  $\text{Fe}^{2+}$ , 2-oxoglutarate,  $\text{O}_2$  and ascorbate for activity. C-P4Hs reside within the lumen of the ER where they catalyze the post-translational hydroxylation of proline residues in -X-Pro-Gly-sequences in collagens and collagen-like proteins. This modification is essential for the formation of stable collagen triple helices. C-P4Hs also play a role in the quality control of collagen synthesis by interacting with and preventing the secretion of unassembled procollagen chains and incorrectly folded procollagens. Collagens are important in healing processes, forming scar and fibrous tissue, but excess collagen accumulation can also be harmful, leading to severe fibrotic diseases, e.g. cirrhosis and pulmonary fibrosis. This has led to attempts to develop drugs that inhibit collagen accumulation. Because of their critical role in collagen synthesis, it has been suggested that C-P4Hs could be potent targets for anti-fibrotic therapy. (for reviews, see Myllyharju & Kivirikko 2001, Koide & Nagata 2005, Myllyharju 2005)

## 2.2.1 Vertebrate collagen prolyl 4-hydroxylases

### *Molecular and in vivo properties*

Three vertebrate C-P4H isoenzymes have been cloned to date from human, mouse, chicken and rat (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). All vertebrate C-P4Hs are tetrameric enzymes consisting of two catalytically active  $\alpha$  subunits and two  $\beta$  subunits. The  $\beta$  subunit is identical to the protein disulfide isomerase (PDI), which has a role as a chaperone in the C-P4H tetramer (Koivu *et al.* 1987, Pihlajaniemi *et al.* 1987). The molecular weight of the C-P4H tetramer is about 240 kDa, those of the  $\alpha$  and  $\beta$  subunits being about 63 kDa and 58 kDa, respectively (for reviews, see Kivirikko & Myllyharju 1998, Myllyharju 2003, 2005). The three  $\alpha$  subunit isoforms,  $\alpha$ (I),  $\alpha$ (II) and  $\alpha$ (III), associate with the same  $\beta$  subunit and form  $[\alpha$ (I)]<sub>2</sub> $\beta$ <sub>2</sub>,  $[\alpha$ (II)]<sub>2</sub> $\beta$ <sub>2</sub> and  $[\alpha$ (III)]<sub>2</sub> $\beta$ <sub>2</sub> tetramers named C-P4Hs I, II and III, respectively (for a review, see Myllyharju 2005). Tetramers containing two different  $\alpha$  subunits are not formed in vertebrates (Annuinen *et al.* 1997).

Expression studies have revealed that C-P4H-I is the main form in most cell types and tissues, whereas the C-P4H-II form predominates in chondrocytes and endothelial cells. C-P4H-III is widely expressed in several adult and fetal tissues but the mRNA level of the  $\alpha$ (III) subunit is much lower than those of the other two. (Helaakoski *et al.* 1994, Annunen *et al.* 1998, Nissi *et al.* 2001, Kukkola *et al.* 2003)

The human  $\alpha$ (I),  $\alpha$ (II) and  $\alpha$ (III) subunits consist of 517, 514 and 525 residues, respectively, and an additional signal peptide of 17 to 21 residues (Helaakoski *et al.* 1989, Annunen *et al.* 1997, Kukkola *et al.* 2003, Van Den Diepstraten *et al.* 2003). The overall amino acid sequence identity between the processed human  $\alpha$ (I) and  $\alpha$ (II) subunits is 64%, while the identity of  $\alpha$ (III) with  $\alpha$ (I) and  $\alpha$ (II) is 35 and 37%, respectively (Annuinen *et al.* 1997, Kukkola *et al.* 2003). The sequence identity is highest in the C-terminal regions where the conserved, catalytically important, amino acids are located (Annuinen *et al.* 1997, Kukkola *et al.* 2003). The peptide substrate binding domain is distinct from the catalytic C-terminal domain and is located in the N-terminal region, and it consists of residues Phe144-Ser244 in the human  $\alpha$ (I) subunit (Fig. 4) (Myllyharju and Kivirikko 1999).

All three  $\alpha$  subunits have five conserved cysteine residues, the  $\alpha$ (II) and  $\alpha$ (III) subunits each having an additional non-conserved cysteine (Annunen *et al.* 1997, Kukkola *et al.* 2003). C-P4Hs have no interchain disulfide bonds between the subunits (Nietfeld *et al.* 1981, Kukkola *et al.* 2003) but site-directed mutagenesis studies of the  $\alpha$ (I) subunit indicate that intrachain disulfide bonds between the second and third and between the fourth and fifth conserved cysteines are essential for the assembly of the  $\alpha_2\beta_2$  tetramer (John & Bulleid 1994, Lamberg *et al.* 1995). All three  $\alpha$  subunits have two N-glycosylation sites and, at least in the case of type I and type III C-P4Hs, glycosylation has no role in the assembly of the enzyme tetramer or in catalytic activity (Lamberg *et al.* 1995, Kukkola *et al.* 2003).

The genes of the human  $\alpha$ (I),  $\alpha$ (II) and  $\alpha$ (III) subunits are located on chromosomes 10, 5 and 11, respectively, consisting of 16, 16 and 13 exons. The genes are very alike in size and have similar exon-intron structures. Exons 9 and 10 in the  $\alpha$ (I) gene and exons 12a and 12b in the  $\alpha$ (II) gene are subjected to alternative splicing. There is no evidence of alternative splicing in the case of the  $\alpha$ (III) gene. (Pajunen *et al.* 1989, Helaakoski *et al.* 1994, Nokelainen *et al.* 2001, Kukkola *et al.* 2003)

PDI, which acts as the  $\beta$  subunit in all vertebrate C-P4Hs, is a multifunctional enzyme that catalyzes the formation, breakage and rearrangement of disulfide bonds during protein folding in the ER. It has other roles in collagen synthesis as well, such as catalyzing intra- and interchain disulfide bond formation in procollagen and preventing secretion of incorrectly folded procollagen molecules (for reviews, see Kivirikko & Myllyharju 1998, Myllyharju 2005). PDI also functions as a  $\beta$  subunit in the microsomal triglyceride transferase dimer and as a chaperone that assists the folding of many newly translated polypeptides. PDI has been cloned and characterized from many species including animal, plant and yeast. (for reviews, see Kivirikko & Myllyharju 1998, Kivirikko & Pihlajaniemi 1998, Freedman *et al.* 2002, Myllyharju 2003, 2005; Wilkinson & Gilbert 2004, Ellgaard & Ruddock 2005)

As a  $\beta$  subunit in vertebrate C-P4Hs, the main function of PDI is to keep the highly insoluble  $\alpha$  subunits in a non-aggregated, catalytically active form (Vuori *et al.* 1992a). It also mediates the retention of the enzyme tetramer in the lumen of the ER (Vuori *et al.* 1992b). The human PDI polypeptide consists of 491 amino acids and is synthesized with a signal sequence of an additional 17 residues (Pihlajaniemi *et al.* 1987). The polypeptide consists of four domains, *a*, *b*, *b'*, *a'* and a highly acidic C-terminal extension *c*, as well as a 19-amino-acid linker

region between the *b'* and *a'* domains (for reviews see, Freedman *et al.* 2002, Wilkinson & Gilpert 2004, Elgaard & Ruddock 2005). Nuclear magnetic resonance studies have shown that all four domains have a thioredoxin fold (Kemink *et al.* 1996, 1997, 1999, Dijkstra *et al.* 1999, Elgaard & Ruddock 2005) and recent crystallization studies of a full-length PDI from yeast have confirmed this (Tian *et al.* 2006). Interestingly, only the amino acid sequences of the catalytic *a* and *a'* domains show homology with the thioredoxin sequence, each containing the characteristic -Cys-Gly-His-Cys- motif (for reviews, see Freedman *et al.* 2002, Wilkinson & Gilpert 2004, Myllyharju 2005). The *b'* domain contains the major peptide binding site and it is also critical for catalysis of isomerization but not oxidation reactions (for reviews, see Freedman *et al.* 2002, Ellgaard & Ruddock 2005). The minimum requirement for the assembly of an active C-P4H tetramer is fulfilled by the PDI domains *b'* and *a'*, while domains *a* and *b* enhance assembly (Pirneskoski *et al.* 2001). It has been demonstrated, however, that point mutations in the *b'* domain do not inhibit C-P4H tetramer assembly and the binding sites in three PDI domains, *a*, *b'* and *a'*, contribute to efficient C-P4H assembly (Koivunen *et al.* 2005). PDI activity is not required for C-P4H tetramer assembly or activity, as mutations in either or both active sites do not affect the ability of PDI to serve as a  $\beta$  subunit of C-P4H (Vuori *et al.* 1992b).

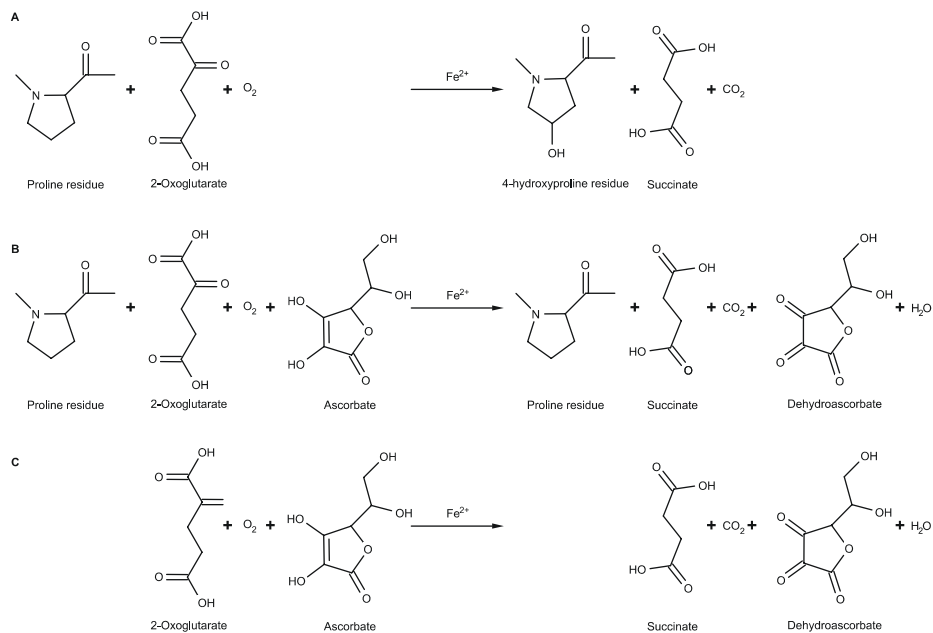
Studies on a knock-out mouse line lacking C-P4H-I activity show that the null mice die around embryonic day 10.5 with an overall developmental delay (Holster *et al.* 2007). The basement membranes of the C-P4H-I null embryos are fragmented and almost completely lack collagen IV, while their fibrillar collagens seem to be quite normal with only slightly increased diameters. The abnormal assembly of type IV collagen is thus most likely the primary cause of death (Holster *et al.* 2007). C-P4H-II knock out mice are born with no obvious phenotypic abnormalities and are viable and fertile (for a review, see Myllyharju 2005). No mutations leading to heritable diseases have been characterized in any of the human C-P4H  $\alpha$  subunit genes (for reviews, see Myllyharju & Kivirikko 2004, Myllyharju 2005).

### *Reaction mechanism and catalytic properties*

C-P4Hs belong to the group of 2-oxoglutarate and non-heme-Fe(II)-dependent dioxygenases and catalyze the formation of 4-hydroxyproline in peptides. The hydroxylation requires  $\text{Fe}^{2+}$ , 2-oxoglutarate,  $\text{O}_2$  and ascorbate (for reviews, see Kivirikko & Myllyharju 1998, Kivirikko & Pihlajaniemi 1998, Myllyharju 2003,

2005). Several studies and theoretical analysis have enabled the elucidation of the reaction mechanism and structural features of the catalytic sites of the C-P4Hs (Myllylä *et al.* 1977, Hanauske-Abel & Günzler 1982, Hanauske-Abel 1991, Lamberg *et al.* 1995, Myllyharju & Kivirikko 1997, Hanauske-Abel & Popowicz 2003).

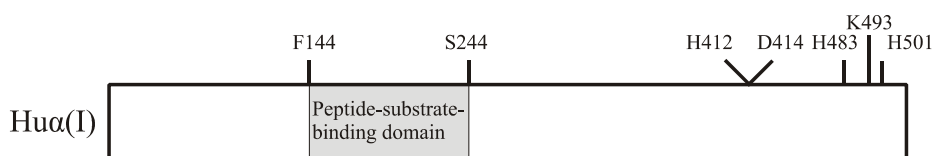
Kinetic studies have demonstrated that the hydroxylation reaction occurs in an ordered manner:  $\text{Fe}^{2+}$  first becomes bound to the enzyme, followed by 2-oxoglutarate,  $\text{O}_2$  and peptide substrate binding (Myllylä *et al.* 1977, Tuderman *et al.* 1977). The order of binding of the last two reactants has remained somewhat uncertain, however, and subsequent studies have shown that the peptide substrate is probably bound before  $\text{O}_2$  (de Jong & Kemp 1984). In the first half of the reaction, the 2-oxoglutarate is decarboxylated and a succinate is formed once one atom of the  $\text{O}_2$  molecule is incorporated into it. This leads to the formation of a highly reactive iron oxo-complex, a ferryl ion, which is the active intermediate and transfers an oxygen atom to the proline in the peptide substrate in the second half of the hydroxylation process (Hanauske-Abel & Günzler 1982, Kivirikko & Myllyharju 1998). Ascorbate is not consumed stoichiometrically, except in uncoupled decarboxylations of 2-oxoglutarate, i.e. decarboxylations without subsequent hydroxylation of the peptide substrate (Fig. 3). C-P4Hs can catalyze a number of reaction cycles in the absence of ascorbate, but as they catalyze uncoupled reaction cycles at a low rate, even in the presence of a saturating concentration of the peptide substrate, they eventually become inactivated (Fig. 3) (for reviews, see Kivirikko & Myllyharju 1998, Kivirikko & Pihlajaniemi 1998, Myllyharju 2005). The main function of ascorbate is thus to reactivate the enzyme as an alternative oxygen acceptor so that it reduces  $\text{Fe}^{3+}\text{-O}^\cdot$  back to the reactive ferryl ion in cases where there are uncoupled reaction cycles (Myllylä *et al.* 1984).



**Fig. 3. The reaction catalyzed by prolyl 4-hydroxylase. 2-oxoglutarate is stoichiometrically decarboxylated during proline hydroxylation (A). In an uncoupled reaction ascorbate is stoichiometrically consumed either in the presence (B) or absence (C) of the substrate.**

The catalytic site of the C-P4H  $\alpha$  subunit appears to comprise a set of separate locations for the binding of cosubstrates and the binding of the peptide substrate (for reviews, see Kivirikko & Myllyharju 1998, Kivirikko & Pihlajaniemi 1998, Myllyharju 2003, 2005). The Fe<sup>2+</sup> is located in a pocket formed by three ligands, His412, Asp414 and His483 in the human  $\alpha$ (I) subunit (Fig. 4) (Lamberg *et al.* 1995, Myllyharju & Kivirikko 1997). The 2-oxoglutarate binding site can be divided into three distinct subsites (Hanuske-Abel & Günzler 1982, Majamaa *et al.* 1984). Subsite I is formed from a positively charged residue, Lys493, in the human  $\alpha$ (I) subunit, which ionically binds the C5 carboxyl group of 2-oxoglutarate (Fig. 4) (Myllyharju & Kivirikko 1997). Subsite II consists of two cis-positioned coordination sites of the enzyme-bound Fe<sup>2+</sup> and is chelated by the C1-C2 moiety, while 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 is an additional critical residue at the

catalytic site (Fig. 4). It directs the orientation of the C1 carboxyl group of 2-oxoglutarate towards the active iron centre, and it accelerates the breakdown of the tetrahedral ferryl intermediate to succinate, CO<sub>2</sub>, and a ferryl ion (Myllyharju & Kivirikko 1997). The ascorbate binding site contains two cis-positioned equatorial coordination sites for the enzyme bound iron and it is thus partially identical to the binding site of 2-oxoglutarate (Majamaa *et al.* 1986). Molecular oxygen is thought to be bound to the Fe<sup>2+</sup> end-on in an axial position, producing the dioxygen unit (Hanauske-Abel & Günzler 1982, Hanauske-Abel & Popowicz 2003). The amino acid residues that have been found to be critical for the binding of Fe<sup>2+</sup> and 2-oxoglutarate to C-P4H-I are conserved in all C-P4Hs studied so far. They are also conserved in other 2-oxoglutarate-dependent dioxygenases, with the exception that the basic residue binding the C5 carboxyl group of 2-oxoglutarate is mostly an arginine instead of a lysine in the C-P4Hs (Myllyharju & Kivirikko 1997, Vranka *et al.* 2004, Clifton *et al.* 2006).



**Fig. 4. Schematic representation of the human C-P4H  $\alpha$ (I) subunit. The catalytically critical residues are indicated above the polypeptide. The peptide binding domain is shown in a grey box.**

The catalytic properties of C-P4Hs I, II and III are very similar, as shown by their comparable  $K_m$  values for cosubstrates and  $K_i$  values for certain inhibitors (Table 1). Nevertheless, some distinct differences exist in their  $K_m$  and  $K_i$  values for peptide substrates and inhibitors (Table 1), suggesting that the peptide binding sites must differ in the three C-P4H isoenzymes. The most striking differences are found in the case of C-P4H-II. It is inhibited by poly(L-proline) only at very high concentrations, while C-P4H-I is inhibited very efficiently and C-P4H-III with an intermediate efficiency. The  $K_m$  value of C-P4H-II for the peptide substrate (Pro-Pro-Gly)<sub>10</sub> is also about fivefold higher than the  $K_m$  values of the type I and III enzymes. (Helaakoski *et al.* 1995, Annunen *et al.* 1997, Kukkola *et al.* 2003)

The C-P4Hs hydroxylate prolines only in a peptide linkage and do not act on free proline. The minimum sequence requirement for hydroxylation is a tripeptide X-Pro-Gly whereas Gly-X-Pro or Pro-Gly-X do not become hydroxylated. Thus only a proline residue preceding the glycine in -X-Y-Gly- sequences in collagens

is hydroxylated. The rate of hydroxylation by vertebrate C-P4Hs is affected by the amino acids present in the X position, a proline in this position resulting in the highest maximal reaction velocity. The chain length of the peptide substrate has a marked effect on the  $K_m$ , which decreases with increasing chain length. The conformation of the peptide substrate has a crucial effect on hydroxylation as a triple-helical conformation totally prevents it. (for reviews, see Kivirikko *et al.* 1992, Myllyharju 2003, 2005)

Studies with individual recombinant peptide binding domains of the human C-P4Hs I and II indicate that the more efficient hydroxylation of longer peptides is most probably due to the higher binding affinity of long peptides for the peptide substrate binding domain rather than properties of the catalytic domain (Hieta *et al.* 2003). The peptide binding domain has been crystallized and was shown to belong to the family of tetratricopeptide repeat (TPR) domains that are involved in many protein-protein interactions (Hieta *et al.* 2003, Pekkala *et al.* 2004).

Several compounds are competitive inhibitors of C-P4Hs and compete with some of the cosubstrates. Many bivalent cations compete with  $Fe^{2+}$ , the most potent being  $Zn^{2+}$ . Several 2-oxoglutarate analogues are effective competitive inhibitors, competing with 2-oxoglutarate, for example, pyridine 2,4-dicarboxylate and pyridine 2,5-dicarboxylate, which have functional groups that can interact at all three subsites of the 2-oxoglutarate binding site (Table 1). Superoxide dismutase-active copper chelates are known to act as competitive inhibitors of C-P4Hs, and compete with  $O_2$ . (for a review, see Myllyharju 2005)

**Table 1.  $K_m$ ,  $K_i$  and  $IC_{50}$  values of vertebrate C-P4Hs I, II and III for the reaction cosubstrates, a peptide substrate and certain competitive inhibitors.**

Cosubstrate, substrate, or inhibitor	Constant	C-P4H-I	C-P4H-II	C-P4H-III
		$\mu\text{M}$	$\mu\text{M}$	$\mu\text{M}$
$\text{Fe}^{2+}$	$K_m$	2 <sup>a</sup>	2 <sup>a</sup>	0.5 <sup>a</sup>
2-Oxoglutarate	$K_m$	20 <sup>a</sup>	22 <sup>a</sup>	20 <sup>a</sup>
Ascorbate	$K_m$	300 <sup>a</sup>	340 <sup>a</sup>	370 <sup>a</sup>
$\text{O}_2$	$K_m$	40 <sup>a</sup>	ND	ND
(Pro-Pro-Gly) <sub>10</sub>	$K_m$	18 <sup>a</sup>	95 <sup>a</sup>	20 <sup>a</sup>
Pyridine 2,4-dicarboxylate	$K_i$	2 <sup>b</sup>	1 <sup>b</sup>	ND
Pyridine 2,5-dicarboxylate	$K_i$	0.8 <sup>a</sup>	ND	ND
3-Hydroxypyridine-2-carbonyl-glycine	$IC_{50}$	0.4 <sup>c</sup>	ND	ND
3,4-Dihydroxybenzoic acid	$K_i$	5 <sup>a</sup>	ND	ND
Poly(L-proline), $M_r$ 5,000-7,000	$IC_{50}$	6 <sup>d</sup>	300 <sup>d</sup>	30 <sup>d</sup>

ND, not determined

<sup>a</sup>Myllyharju 2005, <sup>b</sup>Helaakoski *et al.* 1995, <sup>c</sup>Ivan *et al.* 2002, <sup>d</sup>Kukkola *et al.* 2003

## 2.2.2 Invertebrate collagen prolyl 4-hydroxylases

The nematode *Caenorhabditis elegans* has been widely used to study the outcome of collagen mutations because its exoskeleton is composed predominantly of small collagens (for a review, see Myllyharju & Kivirikko 2004). Four different isoforms of the C-P4H  $\alpha$  subunit have been cloned and characterized from *C. elegans*, called PHY-1, PHY-2, PHY-3 and PHY-4 (Veijola *et al.* 1994, Friedman *et al.* 2000, Hill *et al.* 2000, Winter & Page 2000, Riihimaa *et al.* 2002, Keskiäho *et al.*, unpublished). PHY-1 and PHY-2 are 57% identical to each other and 43-46% identical to the human  $\alpha$ (I) and  $\alpha$ (II) subunits in their amino acid sequence and they play a role in the synthesis of cuticle collagens (Veijola *et al.* 1994, Friedman *et al.* 2000, Hill *et al.* 2000, Winter & Page 2000, Myllyharju *et al.* 2002). PHY-3 and PHY-4 are not involved in cuticle collagen synthesis, but instead PHY-3 hydroxylates collagens in the early embryos, while the precise function of PHY-4 is as yet unknown (Riihimaa *et al.* 2003, Keskiäho *et al.*, unpublished). Three *C. elegans* PDI isoforms have been characterized; PDI-1, PDI-2 and PDI-3. PDI-1 and PDI-2 participate in the assembly of active C-P4Hs, while PDI-3 is more likely to be involved in stabilization of the extracellular matrix or to have a protein chaperone role (Veijola *et al.* 1996, Myllyharju *et al.* 2002, Riihimaa *et al.* 2002, Eschenlauer & Page 2003). The assembly properties of the recombinant *C. elegans* PHY-1 and PHY-2 isoforms differ from those of

the vertebrate C-P4H  $\alpha$  subunits as they can assemble into active PHY-1/PDI-2 and PHY-2/PDI-2 dimers and additionally into a unique mixed tetramer PHY-1/PHY-2/(PDI-2)<sub>2</sub>, which is the main form in *C. elegans* (Veijola *et al.* 1994, Myllyharju *et al.* 2002). Neither PHY-1 nor PHY-2 assembles with PDI-1. Recombinant studies have shown that when PHY-3 is coexpressed with PDI-1, C-P4H activity is generated, but it is unknown whether they form an enzyme complex or whether PDI-1 only assists in the folding of PHY-3 (Riihimaa *et al.* 2002).

*Caenorhabditis briggsae* is a nematode closely related to *C. elegans*. The amino acid identity of their PHY-1 and PHY-2 subunits is 92-97%, and therefore it is surprising that the assembly of active C-P4H forms in them revealed distinct differences. In addition to a PHY-1/PHY-2/(PDI-2)<sub>2</sub> tetramer and a PHY-1/PDI-2 dimer, an active (PHY-2)<sub>2</sub>/(PDI-2)<sub>2</sub> tetramer is formed in *C. briggsae* instead of a PHY-2/PDI-2 dimer. (Winter *et al.* 2007)

A C-P4H  $\alpha$  subunit has also been cloned and characterized from the filarial parasitic nematodes, *Onchocerca volvulus* and *Brugia malayi* (Merriweather *et al.* 2001, Winter *et al.* 2003). The C-P4H of *B. malayi* exists as a very unique  $\alpha_4$  homotetramer that is active and soluble without PDI as a subunit (Winter *et al.* 2003). The cuticle of filarial nematodes is the interface between the parasite and its host, making it a potential target for chemotherapeutic and immunotherapeutic attack. Detailed studies on nematode C-P4Hs may facilitate identification of specific inhibitors of parasite C-P4H enzymes (Friedman *et al.* 2000, Winter & Page 2000, Merriweather *et al.* 2001, Winter *et al.* 2003).

*Drosophila melanogaster* has a family of 19 members which code for C-P4H  $\alpha$  subunit-like polypeptides (Abrams & Andrew 2002). This is surprising because *D. melanogaster* has only four genes coding for collagens and a protein with a collagen-like domain (pericardin) thus indicating that some of the *D. melanogaster* C-P4H  $\alpha$  subunit-like polypeptides may hydroxylate proline residues in proteins other than collagens (for a review, see Myllyharju 2005). Three of the C-P4H  $\alpha$  subunit-like polypeptides have been characterized in detail (Annunen *et al.* 1999, Abrams *et al.* 2006). The kinetic properties of one of these shows that its substrate specificity differs from the human C-P4Hs as the peptide substrate (Pro-Pro-Gly)<sub>10</sub> is hydroxylated very inefficiently (Annunen *et al.* 1999). Furthermore, the peptide binding region of this *D. melanogaster* polypeptide shows a relatively low identity with the corresponding domain of human C-P4H  $\alpha$  subunits (Annunen *et al.* 1999). The two other characterized C-P4H  $\alpha$  subunit-like polypeptides have been shown to be salivary gland-specific and are necessary for

normal salivary gland secretion and the maintenance of a uniform salivary gland lumen (Abrams *et al.* 2006). These two isoforms are likely to hydroxylate some secreted or transmembrane proteins other than collagen, as the embryonic salivary glands do not express collagen.

## **2.3 Plant and viral prolyl 4-hydroxylases**

### **2.3.1 Higher plant and algal prolyl 4-hydroxylases**

Plant P4Hs have been partially purified and characterized from the higher plant sources *Vinca rosea* (Tanaka *et al.* 1980), the French bean *Phaseolus vulgaris* (Bolwell *et al.* 1985, Wojtaszek *et al.* 1999), and the green algae *Chlamydomonas reinhardtii* (Kaska *et al.* 1987) and *Volvox carteri* (Kaska *et al.* 1988). These have been shown to be soluble monomers. Four plant genes encoding P4H polypeptides have been cloned and characterized, from *Arabidopsis thaliana* (Hieta & Myllyharju 2002, original article II of this study), *Nicotiana tabacum* (Yuasa *et al.* 2005) and *C. reinhardtii* (Keskiaho *et al.* 2007).

The genome of *A. thaliana* encodes at least six P4H-like polypeptides which show 21-27% identity to the catalytic C-terminal regions of the human C-P4H  $\alpha$ (I) and  $\alpha$ (II) subunits (Hieta & Myllyharju 2002). One of these cDNAs, encoding a 283-residue soluble monomer, At-P4H-1, has been cloned and characterized (Hieta & Myllyharju 2002). At-P4H-1 hydroxylates poly(L-proline) efficiently, the  $K_m$  values for poly(L-proline)  $M_r$  5,000 and 20,000 being 2 and 0.2  $\mu$ M, respectively (Hieta & Myllyharju 2002). *A. thaliana* AGP and extensin sequences are also efficiently hydroxylated, the  $K_m$  values varying between 10 and 40  $\mu$ M (Hieta & Myllyharju 2002). Although collagen does not exist in plants, surprisingly, the recombinant At-P4H-1 also effectively hydroxylated collagen-like sequences, the  $K_m$  value for (Pro-Pro-Gly)<sub>10</sub> being 60  $\mu$ M, as well as peptides representing the N- and C-terminal hydroxylation sites in the human HIF-1 $\alpha$  subunit, with  $K_m$  values of 100 and 50  $\mu$ M, respectively (Hieta & Myllyharju 2002). This indicates that a poly(L-proline) type II helix conformation is not an essential requirement for plant P4H hydroxylation (Hieta & Myllyharju 2002), as has been previously suggested (for a review, see Kivirikko *et al.* 1992).

The *C. reinhardtii* genome encodes 10 putative P4H-like polypeptides and one of these, Cr-P4H-1, which consists of 253 amino acids, has been cloned and characterized. The sequence of Cr-P4H-1 is 26% identical to that of the catalytic

C-terminal region of the human C-P4H  $\alpha$ (I) subunit, the identity with the At-P4H-1 sequence being 40%. The Cr-P4H-1 efficiently hydroxylates poly(L-proline), the  $K_m$  values for poly(L-proline)  $M_r$  5,000 and 20,000 being 125 and 2  $\mu$ M, respectively. Synthetic peptides representing the HRGP GP1 found in the *C. reinhardtii* outer cell wall were also efficiently hydroxylated. Suppression of the gene encoding Cr-P4H-1 by RNA interference led to defective cell walls proving that 4-hydroxyproline has an important role in the assembly of the algal cell wall. (Keskiäho *et al.* 2007)

The plant P4Hs differ in their substrate requirements when compared with animal C-P4Hs as they efficiently hydroxylate poly(L-proline) which is not hydroxylated by C-P4Hs, but instead is an effective inhibitor of some of them (Table 2) (for reviews, see Kivirikko *et al.* 1992, Kivirikko & Myllyharju 1998). They also hydroxylate collagens, although less efficiently than the animal C-P4Hs (compare Table 1 and 2). The cosubstrate requirements of plant P4Hs are the same as those of animal C-P4Hs, the  $K_m$  values being highly similar (compare Table 1 and 2). The most striking difference in cosubstrate requirements between plant P4Hs and human C-P4H-I is the markedly higher  $K_m$  value, 130 to 250  $\mu$ M, of the former for 2-oxoglutarate (Hieta & Myllyharju 2002, Keskiäho *et al.* 2007). As plant P4Hs have been shown to be soluble monomers they are regarded as promising targets for structural studies. Information about the structure of their catalytic site could serve as a model for that of the animal C-P4Hs.

**Table 2.  $K_m$  values of At-P4H-1 and Cr-P4H-1 for reaction cosubstrates and substrates.**

Cosubstrate, substrate	Constant	At-P4H-1 <sup>a</sup> $\mu$ M	Cr-P4H-1 <sup>b</sup> $\mu$ M
Fe <sup>2+</sup>	$K_m$	16	30
2-Oxoglutarate	$K_m$	130	250
Ascorbate	$K_m$	300	20
Poly(L-proline), $M_r$ 5,000-10,000	$K_m$	2	140
(Pro-Pro-Gly) <sub>10</sub>	$K_m$	60	>1500

<sup>a</sup>Hieta & Myllyharju 2002, <sup>b</sup>Keskiäho *et al.* 2007

### 2.3.2 *Paramecium bursaria Chlorella virus-1 prolyl 4-hydroxylase*

Viruses and bacteria have been shown to have proline-rich polypeptides as well as short collagen-like sequences (Smith *et al.* 1998, Eriksson *et al.* 1999, Xu *et al.* 2002, Rasmussen *et al.* 2003). Bacteria are not known to possess peptidyl P4H

activity, however, but a viral P4H has been cloned from a eukaryotic algal virus, *Paramecium bursaria Chlorella virus-1* (PBCV-1) (Eriksson *et al.* 1999). It is a 242-residue polypeptide, which shows a distinct sequence similarity of 15-23% to the corresponding C-terminal residues of various catalytic  $\alpha$  subunits of animal C-P4Hs (Eriksson *et al.* 1999). The recombinant PBCV-1 P4H is a monomer and hydroxylates poly(L-proline) and several synthetic peptides corresponding to proline-rich repeats coded by the viral genome (Eriksson *et al.* 1999). Like At-P4H-1, the PBCV-1 P4H also hydroxylated (Pro-Pro-Gly)<sub>10</sub>, but inefficiently, the  $K_m$  value being about 3 mM (Eriksson *et al.* 1999).

## 2.4 HIF prolyl 4-hydroxylases

As described in chapter 2.1.2, under normoxic conditions the HIF- $\alpha$  subunit is targeted for degradation by an ubiquitin-ligase complex which recognizes hydroxylated proline residue(s) (Ivan *et al.* 2001, Jaakkola *et al.* 2001). The hydroxylation is performed in vertebrates by a family of HIF-P4Hs, HIF-P4H-1, 2 and 3 (also called EglN2, 1, 3, PHD1, 2, 3, and HPH2, 1, 3, respectively) (Bruick & McKnight 2001, Epstein *et al.* 2001, Ivan *et al.* 2002). HIF-P4Hs have also been characterized from the invertebrates *C. elegans* and *D. melanogaster*, both of which only have a single HIF-P4H (Bruick & McKnight 2001, Epstein *et al.* 2001).

### *Molecular and in vivo properties*

Like the C-P4Hs, HIF-P4Hs belong to the 2-oxoglutarate dependent dioxygenase family (for reviews, see Schofield & Zhang 1999, Schofield and Ratcliffe 2005). The reaction mechanism is thus similar to that of the C-P4Hs (see section 2.2.1). The catalytically critical residues, two histidines and one aspartate that bind the iron, and an arginine that binds the C5 carboxyl group of 2-oxoglutarate, are located within the C-terminal region of the HIF-P4H polypeptides and are conserved between all three isoenzymes (Bruick and McKnight 2001, Epstein *et al.* 2001, Ivan *et al.* 2002). An arginine in HIF-P4Hs replaces the lysine in C-P4Hs as the binding site for the C5 carboxyl group of 2-oxoglutarate. HIF-P4Hs share homology in the C-terminal catalytic part but have marked differences in their N-terminal sequences and do not have notable overall sequence homology to the catalytic subunits of C-P4Hs (Bruick & McKnight 2001, Epstein *et al.* 2001, Ivan *et al.* 2002). The three-dimensional structure of an N-terminally truncated

HIF-P4H-2 has been solved and shown to resemble those of other 2-oxoglutarate dependent dioxygenases in that it has a double-stranded  $\beta$ -helix “jelly-roll” core fold (McDonough *et al.* 2006).

The human HIF-P4H-1, 2 and 3 polypeptides consist of 407, 426 and 239 amino acids, respectively (Taylor 2001, Bruick & McKnight 2001, Epstein *et al.* 2001, Ivan *et al.* 2002). The genes of HIF-P4H-1, 2 and 3 are located on chromosomes 19, 1 and 14, respectively (Taylor 2001). At least two catalytically inactive splicing variants of HIF-P4H-2 and one of HIF-P4H-3 exist, while another splicing variant of HIF-P4H-3 is at least partially active (Hirsilä *et al.* 2003, Cervera *et al.* 2006). The HIF-P4H-2 and 3 genes, but not that of HIF-P4H-1, have been found to be hypoxia-inducible (Cioffi *et al.* 2003, Aprelikova *et al.* 2004). HIF-P4H-2 is the most abundantly expressed isoenzyme and under normal circumstances is the main form responsible for the hydroxylation of HIF- $\alpha$ , while HIF-P4H-1 and 3 become important after prolonged periods of hypoxia (Berra *et al.* 2003, Appelhoff *et al.* 2004). Knock-out mouse studies suggest that HIF-P4H-2 is the most important isoform as HIF-P4H-2 null mice are embryonically lethal, while HIF-P4H-1 and 3 knock-out mice are apparently normal (Takeda *et al.* 2006). HIF-P4Hs are cytoplasmic and nuclear proteins, HIF-P4H-1 being located exclusively in the nucleus, HIF-P4H-2 mainly in the cytoplasm and HIF-P4H-3 being found in both compartments (Metzen *et al.* 2003a). mRNA expression studies have revealed that although all three HIF-P4Hs are widely expressed in many tissues, they exhibit tissue specific expression patterns (Lieb *et al.* 2002).

### *Catalytic properties*

Hydroxylation of HIF- $\alpha$ s occurs in conserved -Leu-X-X-Leu-Ala-Pro- sequences in two sites in HIF-1 $\alpha$  (Pro<sup>402</sup> and Pro<sup>564</sup>) and HIF-2 $\alpha$ , and in one site in HIF-3 $\alpha$  (Bruick & McKnight 2001, Masson *et al.* 2001). However, it has been shown that both leucines can be replaced with many other residues with essentially no effect on the hydroxylation (Huang J *et al.* 2002, Li *et al.* 2004), alanine being the only relatively, but not absolutely, strict requirement in addition to the proline itself (Li *et al.* 2004). The C-P4Hs do not hydroxylate proline residues in the HIF- $\alpha$  subunits (Jaakkola *et al.* 2001) as they have a distinct substrate specificity and are located in a different cellular compartment (for a review, see Myllyharju 2005).

The catalytic and inhibition properties of the three HIF-P4Hs are very similar, however, HIF-P4H-3 has some distinct features (Table 3) (Hirsilä *et al.* 2003, 2005; Koivunen *et al.* 2006, 2007). The HIF-P4Hs, especially HIF-P4H-3,

hydroxylate the C-terminal hydroxylation site (Pro<sup>564</sup> in HIF-1 $\alpha$ ) much more efficiently than the N-terminal site (Pro<sup>402</sup> in HIF-1 $\alpha$ ) (Hirsilä *et al.* 2003, 2005, Chan *et al.* 2005, Koivunen *et al.* 2006). HIF-P4Hs resemble C-P4Hs in that the length of the peptide substrate has a marked effect on the hydroxylation efficiency, the  $K_m$  values for recombinant HIF-1 $\alpha$  and HIF-2 $\alpha$  ODDD of 248 and 215 residues, respectively, being markedly lower than for synthetic 19-20 and 35-residue peptides (Koivunen *et al.* 2006). Studies in cultured cells have indicated that the actions of the HIF-P4Hs on different HIF- $\alpha$  isoforms are not equivalent (Appelhoff *et al.* 2004). As recent kinetic studies have not shown significant differences in the  $K_m$  values of the HIF-P4Hs for the different HIF- $\alpha$  ODDDs (Koivunen *et al.* 2006), it is likely that the different actions of HIF-P4Hs within cells reflect differences in their abundance and cellular location.

Although the HIF-P4Hs and C-P4Hs have an identical reaction mechanism, their  $K_m$  values for cosubstrates differ significantly (compare Table 1 and 3). The HIF-P4Hs bind iron and 2-oxoglutarate more tightly than the C-P4Hs (Koivunen *et al.* 2007). The  $K_m$  values of HIF-P4Hs for iron are 20 to 65 times lower and those of HIF-P4H-1 and 2 for 2-oxoglutarate are 10 to 20 times lower than the corresponding values for C-P4Hs. The 2-oxoglutarate binding properties of HIF-P4H-3 are intermediate between those of the HIF-P4H-1 and 2 and the C-P4Hs. All HIF-P4Hs have similar and very high  $K_m$  values for O<sub>2</sub>, 230-250  $\mu$ M, as determined using a 19-residue synthetic peptide as a substrate (Hirsilä *et al.* 2003). Their affinity for oxygen is affected by the length of the substrate, however. This was shown recently when a full-length ODDD was used as a substrate and the  $K_m$  of HIF-P4H-2 for oxygen decreased from 250 to 100  $\mu$ M (Koivunen *et al.* 2006). Even this value is still much higher than the O<sub>2</sub> concentration in tissues and is thus rate limiting for HIF-P4H-2 activity under physiological conditions. This supports the theory that HIF-P4Hs function as critical oxygen sensors.

To date, HIF- $\alpha$  subunits are the only known substrates for HIF-P4Hs, but there is indirect evidence that other substrates are likely to exist. Likely candidates are, for example, the iron regulatory protein-2, the large subunit of RNA polymerase II, I $\kappa$ B kinase- $\beta$  and activating transcription factor 4 (Kuznetsova *et al.* 2003, Wang & Pantopoulos 2005, Cummins *et al.* 2006, Köditz *et al.* 2007).

Several synthetic 2-oxoglutarate analogues (Table 3), as well as the citric acid cycle intermediates fumarate and succinate, have been shown to be effective competitive inhibitors of all HIF-P4Hs (Ivan *et al.* 2002, Hirsilä *et al.* 2003, Isaacs *et al.* 2005, Selak *et al.* 2005, Hewitson *et al.* 2007, Koivunen *et al.* 2007).

Reactive oxygen species and nitric oxide have also been reported to inhibit HIF-P4Hs, however, the precise mechanism of the inhibition remains to be examined (Metzen *et al.* 2003b). In contrast to C-P4Hs, HIF-P4H-1 and 2 are inhibited very inefficiently by most metals, while HIF-P4H-3 is inhibited more effectively (Hirsilä *et al.* 2005). Cobalt and nickel are known to stabilize HIF- $\alpha$  effectively, however, but this is likely to be due to a more complicated mechanism than simple competitive inhibition (Hirsilä *et al.* 2005). Stabilization of HIF- $\alpha$  polypeptides by small molecule inhibitors of HIF-P4Hs is believed to be therapeutically beneficial in diseases characterized by acute or chronic ischemia or severe anaemia (for reviews, see Kaelin 2005a, Bruegge *et al.* 2007). There seem to be differences in the inhibitory properties of HIF-P4Hs and C-P4Hs, and even between individual HIF-P4H isoenzymes (compare Table 1 and 3). Thus, it should be possible to develop specific inhibitors for both enzyme groups.

**Table 3.  $K_m$  and  $K_i$  values of HIF-P4H-1, 2 and 3 for reaction cosubstrates, substrates and certain inhibitors.**

Cosubstrate, substrate, or inhibitor	Constant	HIF-P4H-1 $\mu\text{M}$	HIF-P4H-2 $\mu\text{M}$	HIF-P4H-3 $\mu\text{M}$
$\text{Fe}^{2+}$	$K_m$	0.03 <sup>a</sup>	0.03 <sup>a</sup>	0.1 <sup>a</sup>
2-Oxoglutarate	$K_m$	2 <sup>b</sup>	1 <sup>b</sup>	12 <sup>b</sup>
Ascorbate	$K_m$	170 <sup>c</sup>	180 <sup>c</sup>	140 <sup>c</sup>
$\text{O}_2$ (19-residue synthetic peptide as a substrate)	$K_m$	230 <sup>c</sup>	250 <sup>c</sup>	230 <sup>c</sup>
(HIF-1 $\alpha$ ODDD as a substrate)			100 <sup>d</sup>	
HIF-1 $\alpha$ ODDD	$K_m$	0.01-0.02 <sup>d</sup>	0.14 <sup>d</sup>	0.07 <sup>d</sup>
HIF-2 $\alpha$ ODDD	$K_m$	0.01-0.02 <sup>d</sup>	0.06 <sup>d</sup>	0.1 <sup>d</sup>
Pyridine 2,4-dicarboxylate	$K_i$	40 <sup>c</sup>	7 <sup>c</sup>	8 <sup>c</sup>
Pyridine 2,5-dicarboxylate	$K_i$	>300 <sup>c</sup>	>300 <sup>c</sup>	>300 <sup>c</sup>
3-Hydroxypyridine-2-carbonyl-glycine	$K_i$	15 <sup>c</sup>	2 <sup>c</sup>	1 <sup>c</sup>
3,4-Dihydroxybenzoic acid	$K_i$	>300 <sup>c</sup>	>300 <sup>c</sup>	>300 <sup>c</sup>

<sup>a</sup>Hirsilä *et al.* 2005, <sup>b</sup>Koivunen *et al.* 2007, <sup>c</sup>Hirsilä *et al.* 2003, <sup>d</sup>Koivunen *et al.* 2006

## 2.5 Prolyl 3-hydroxylases

3-Hydroxyproline is a rare amino acid in proteins. To date, 3-hydroxyproline residues are only known to exist in vertebrates in collagens where they are found in the sequence -Gly-3Hyp-4Hyp-Gly- (Fietzek *et al.* 1972, Gryder *et al.* 1975, Tryggvason *et al.* 1977). The extent of 3-hydroxylation varies in different types of

collagens and is most abundant in type IV collagen where the 3-hydroxyproline content mostly varies between 5 to 10 residues, but can be up to 16 residues per 1,000 amino acids, depending on the tissue source (Kefalides 1973, Gryder *et al.* 1975, Kresina & Miller 1979). In contrast, type I and X collagens have only one 3-hydroxyproline residue per  $\alpha$  chain (Fietzek *et al.* 1972, Bos *et al.* 1999) and type II collagen has about two residues per  $\alpha$  chain (Miller *et al.* 1976). About 11% of the total hydroxyproline in type IV collagen is 3-hydroxyproline, and this amount is 10 times higher than in the fibril forming collagens (Kefalides 1973).

Prolyl 3-hydroxylase (P3H, EC 1.14.11.7) catalyzes the hydroxylation of proline residues in the -Gly-Pro-4Hyp-Gly- sequences in vertebrates. The formation of 3-hydroxyproline is dependent on the presence of 4-hydroxyproline in the sequence (Risteli *et al.* 1977, Tryggvason *et al.* 1977). P3H was purified and partially characterized about 30 years ago (Risteli *et al.* 1977, Tryggvason *et al.* 1979), but it was only cloned from chick in 2004 (Vranka *et al.* 2004). Three human genes coding for P3H isoenzymes, P3H1, 2 and 3, located in chromosomes 1, 3 and 12, respectively, are now known (Järnum *et al.* 2004, Vranka *et al.* 2004). The human P3H1, 2 and 3 consist of 736, 708 and 726 residues, respectively, including a presumptive signal peptide in their N-terminus (Järnum *et al.* 2004, Vranka *et al.* 2004). The isoenzymes share 38 to 46% sequence identity with each other (Vranka *et al.* 2004). All isoenzymes contain an ER retrieval signal, a -Lys-Asp-Glu-Leu in the case of P3H1 and P3H2 and an -Arg-Glu-Glu-Leu in the case of P3H3 (Vranka *et al.* 2004). The P3H1 and P3H2 polypeptides are also predicted to contain TPR domains (Järnum *et al.* 2004). These domains are known to be involved in protein-protein interactions and have been shown to compose the 3-dimensional structure of the peptide substrate binding domain of human C-P4H-I (Pekkala *et al.* 2004). The TPR domain structure is also found in cartilage associated protein (CRTAP), which shares similarity with the P3Hs, excluding their catalytic residues (Morello *et al.* 2006). The P3Hs share the conserved catalytic residues of the P4Hs and require the same cosubstrates (Risteli *et al.* 1977, Tryggvason *et al.* 1979, Vranka *et al.* 2004), thus belonging to the 2-oxoglutarate dependent dioxygenase family (Kivirikko & Pihlajaniemi 1998, Aravind & Koonin 2001).

P3H1 was originally described as a putative basement membrane associated proteoglycan, Leprecan (Wassenhove-McCarthy & McCarthy 1999), and subsequently as a potential growth suppressor gene, GROS1, on chromosome 1 (Kaul *et al.* 2000). Leprecan was shown to have an ER-Golgi staining pattern in cultured cells (Wassenhove-McCarthy & McCarthy 1999). P3H1 has been

purified from a chick embryo rough ER extract and partial sequencing of the polypeptide led to cloning of the chick P3H1 (Vranka *et al.* 2004). The purified chick P3H1 was shown to have P3H activity on a full-length prolyl 4-hydroxylated procollagen substrate (Vranka *et al.* 2004). The P3H1 was present in a complex of proteins that specifically bound to non-triple-helical collagen. CRTAP and cyclophilin B were tightly bound to P3H1 but neither of these was required for P3H1 activity *in vitro* (Vranka *et al.* 2004, Morello *et al.* 2006). As 3-hydroxyproline is most abundant in type IV collagen of the basement membranes it was surprising that P3H1 was localized in tissues that express fibrillar collagens, and expression in tissues where basement membrane collagens predominate, e.g. kidney and liver, was limited, if present at all (Vranka *et al.* 2004). P3H2 was first identified as a Leprecan-like 1 protein and cloned from a liposarcoma cell line cDNA (Järnum *et al.* 2004). The strongest mRNA expression was seen in kidney and placenta (Järnum *et al.* 2004). The P3H2 polypeptide was located in the ER-Golgi network and expressed in several human tissues such as liver, pancreas, salivary gland, heart and skeletal muscle (Järnum *et al.* 2004).

The function of 3-hydroxyproline residues in collagens is not well known and their biological role and importance remained elusive until not long ago. Recently, biophysical studies were carried out on synthetic polypeptides that demonstrated that the presence of 3-hydroxyproline in both the X and Y positions of a collagen-like peptide destabilized the triple-helical structure (Jenkins *et al.* 2003, Mizuno *et al.* 2004). This effect is totally opposite to that of the 4-hydroxyproline residues which markedly increase the stability of the triple helix. The destabilizing effect is weaker when the 3-hydroxyproline is in its natural X position and stronger when present in the unnatural Y position (Jenkins *et al.* 2003). This destabilizing effect brought up the theory that 3-hydroxyproline residues could serve to modulate the local stability of triple helices, allowing specific regions to have lower stability. This might be necessary for the assembly of certain supramolecular structures, e.g. the meshwork structure formed by type IV collagen in basement membranes (Jenkins *et al.* 2003, Mizuno *et al.* 2004). The collagen triple helices in basement membranes interact with each other, as well as with other biomolecules, in a more varied and complex manner than those present in collagen fibrils (Jenkins *et al.* 2003). The latest structural studies show, however, that the presence of 3-hydroxyproline does not lead to large structural alterations in the collagen triple helix and is more likely to be involved in its protein interactions (Schumacher *et al.* 2006).

Even though the exact function of 3-hydroxyproline residues in collagens is unclear, studies on gene-modified mice and patient samples indicate that the prolyl 3-hydroxylation of type I collagen is crucial for normal bone formation *in vivo* (Barnes *et al.* 2006, Morello *et al.* 2006, Cabral *et al.* 2007). CRTAP has been shown to form a tight complex with P3H1 (Vranka *et al.* 2004, Morello *et al.* 2006) and is required for efficient prolyl 3-hydroxylation of fibrillar collagens *in vivo* (Morello *et al.* 2006). Crtap null mice are characterized by a total loss of 3-hydroxyproline in fibrillar type I and II collagens. The lack of CRTAP function leads to changes in collagen modifications, intracellular trafficking, and extracellular fibril assembly (Morello *et al.* 2006). These abnormalities result in functional defects in osteoid production and the mineralization rate (Morello *et al.* 2006). In humans, CRTAP mutations have been shown to cause a recessive type VII osteogenesis imperfecta (OI) ranging from neonatal lethal cases to a milder phenotype, depending on the type of mutation (Morello *et al.* 2006, Barnes *et al.* 2006). CRTAP has also been shown to have the same chromosomal location as a locus for a moderately severe recessive type VII OI (Tonachini *et al.* 1999, Labuda *et al.* 2002). The majority of OI cases are caused by mutations in the type I collagen genes, COL1A1 and COL1A2 (Rauch & Glorieux 2004). In about 10-15% of OI cases, no collagen mutation is found and a preliminary estimation is that CRTAP defects cause 2 to 3% of all lethal OI cases (Barnes *et al.* 2006). A new recessive human bone disorder resulting from null P3H1 alleles has also been reported (Cabral *et al.* 2007). The lack of P3H1 results in excess lysyl hydroxylation and subsequent over-modification of the type I collagen, which leads to a phenotype that overlaps with lethal or severe OI but has some distinctive features (Cabral *et al.* 2007).

The presence of 3-hydroxyproline in collagen seems to be restricted to vertebrates as it is not found in nematode collagens, although it is abundant in several other proteins, e.g. cathepsin-L-like proteinases and a Kunitz-type molecule in the trematode *Fasciola hepatica* (Wijffels *et al.* 1994, Bozas *et al.* 1995, Bozas & Spithill 1996). The 3-hydroxyproline residues of the *F. hepatica* proteins exist in sequences that have no consensus between each other and no similarity to a collagen sequence (Wijffels *et al.* 1994, Bozas & Spithill 1996). It thus seems likely that the P3H present in this parasite is quite different from the collagen P3Hs of vertebrates (Bozas *et al.* 1995, Wijffels *et al.* 1994).

### 3 Outlines of the present research

4-Hydroxyproline residues were thought to be found only in collagens and collagen-like proteins until recently when it was discovered that one or two 4-hydroxyproline residues are also present in HIF- $\alpha$  polypeptides in normoxia. This led to the identification and characterization of a new P4H family, HIF-P4Hs. Thus, two P4H families are currently known: C-P4Hs that reside in the ER, and HIF-P4Hs that are located either in the nucleus or cytoplasm. Both P4H families belong to the 2-oxoglutarate dioxygenases and require  $\text{Fe}^{2+}$ , 2-oxoglutarate,  $\text{O}_2$  and ascorbate in their reaction.

The discovery of HIF-P4Hs has highlighted the possibility that 4-hydroxyproline residues may yet be found in other proteins and that other groups of P4Hs may also exist. Our database searches revealed a novel human C-P4H-like polypeptide, the sequence of which did not show homology to the peptide substrate binding domain of the C-P4Hs, however. We set out i) to clone and characterize the novel C-P4H-like polypeptide. As it contained a transmembrane domain and was thus difficult to purify in an active form and as all plant P4Hs characterized previously have been shown to be soluble monomers, we decided ii) to clone and characterize a second *A. thaliana* P4H that shared a marked homology with the novel human P4H-like polypeptide at its putative peptide substrate binding region.

When this thesis work was initiated, only very limited data was available on P3Hs, which are also collagen modifying enzymes that belong to the 2-oxoglutarate dioxygenase superfamily. Their importance in collagen synthesis was largely unknown. Thus we set out iii) to clone the three human P3H isoenzymes, express them as recombinant proteins, determine their catalytic properties and study their expression in different tissues at mRNA and protein level.



## 4 Materials and methods

The materials and methods used in this thesis are summarized in Tables 4-6 below. They have been described in detail, together with their references, in original articles I-III.

**Table 4. Constructs for recombinant expression of P4H-TM, At-P4H-2 and the three P3H polypeptides.**

Construct	Description	Used in
Baculovirus expression constructs		
P4H-TM in pVL1392	aas 1-502 of P4H-TM, full-length	I
P4H-TM+PDI+His6 tag in pVL1392	aas 88-502 of P4H-TM with the PDI signal peptide followed by a His6 tag in the N-terminus, lacks the transmembrane domain	I
P4H-TM+GP67+His6 tag in pAcGP67B	aas 88-502 of P4H-TM with the GP67 signal peptide followed by a His6 tag in the N-terminus, lacks the transmembrane domain	I
At-P4H-2 in pVL 1392	aas 1-299 of At-P4H-2, full-length	II
P3H1 in pVL 1392	aas 1-736 of P3H1, full-length	III
P3H2 in pVL 1392	aas 1-708 of P3H2, full-length	III
P3H3 in pVL 1392	aas 1-736 of P3H3, full-length	III
Mammalian expression construct		
P4H-TM in pcDNA3.1(-)Zeo	aas 1-502 of P4H-TM, full-length	I

**Table 5. Antibodies.**

Antibody	Description of the antigen	Type	Used in
P4H-TM	aas 88-502 of human P4H-TM, lacks the transmembrane domain	Polyclonal, rabbit	I
P3H1	Mouse, peptide PEEVIPKRLQEKGKSE	Polyclonal, rabbit	III
P3H2	Mouse, peptide MGKKSPPKIGRDLREG	Polyclonal, rabbit	III

**Table 6. Methods.**

Level	Method	Used in
DNA	Cloning techniques	I, II, III
	PCR	I, II, III
	Rapid amplification of cDNA ends (RACE)	I
	Site-directed mutagenesis	III
RNA	Northern Blot analysis	I
	RT-PCR	I
	siRNA	I
Protein	Expression and analysis of recombinant proteins in mammalian cells	I
	Expression and analysis of recombinant proteins in insect cells	I, II, III
	SDS-PAGE and Western blotting	I, II, III
	Enzyme activity assays	I, II, III
	Metal chelate affinity chromatography	I
	Gel filtration	I
	N-terminal sequencing	I, II, III
	N glycosidase F treatment	I, II
	Circular dichroism spectroscopy	I
	In vitro transcription translation	I
	HPLC	II
Cell and tissue	Cell culture	I, II, III
	Immunofluorescence staining	I, III
	Immunoelectron microscopy (EM)	I, III

## 5 Results

### 5.1 Cloning and characterization of a novel ER transmembrane human prolyl 4-hydroxylase, P4H-TM (I)

#### 5.1.1 Cloning and expression of recombinant P4H-TM

A gene bank search identified a novel P4H-like polypeptide of 502 residues that, unlike any other P4H characterized so far, contained a predicted transmembrane domain spanning residues 59-82 and no signal peptide (Figure 1 in I). The novel polypeptide was named P4H-TM and database searches indicated that it was also found in other vertebrates but not in invertebrates, the zebrafish homologue being 51% identical with the human P4H-TM.

The overall sequence identity of the full-length P4H-TM with the catalytic  $\alpha$  subunits of the three human C-P4Hs and HIF-P4Hs is relatively low, 14-15% and 10-13%, respectively (Figure 1 in I). The identity is higher within the catalytically important C-terminal regions, however, the P4H-TM residues 314-502 being 26-28% and 13-15% identical to the corresponding residues in the C-P4H  $\alpha$  subunits and the HIF-P4Hs, respectively. Notably, the residue binding the C-5 carboxyl group of 2-oxoglutarate is a lysine in P4H-TM (Figure 1 in I) as in the C-P4Hs, while it is an arginine in the HIF-P4Hs (Myllyharju & Kivirikko 1997, Bruick and McKnight 2001, Epstein *et al.* 2001, Ivan *et al.* 2002, Clifton *et al.* 2006). The catalytic region of P4H-TM is thus more closely related to those of the C-P4Hs than the HIF-P4Hs. P4H-TM lacked sequence similarity with the C-P4Hs at their peptide substrate binding domain, however (Figure 1 in I). The C terminus of P4H-TM has the sequence -Arg-Val-Glu-Leu, which according to novel unpublished data functions as an ER retention signal.

To study the properties of recombinant P4H-TM, untagged full-length P4H-TM was expressed as a recombinant protein in insect cells. The cells were harvested 72 h after infection and homogenized in a buffer containing Triton X-100 and centrifuged, and the remaining pellets were further solubilized in 1% SDS. The Triton X-100 and SDS-soluble fractions were analyzed by SDS-PAGE under reducing conditions followed by Coomassie Blue staining or Western blotting (Figure 3A in I) with a P4H-TM antibody. Most of the full-length P4H-TM polypeptide remained in the SDS extract, but significant amounts were also soluble in the Triton X-100 buffer (Figure 3A in I). The recombinant P4H-TM

was found in two forms (Figure 3A in I) that according to N-terminal sequencing started from Met1 and Asp88. This indicates that some of the full-length polypeptides were cleaved after the transmembrane domain. The cleavage was also found to occur in mammalian cells as analysis of samples from human embryonic kidney HEK293 cells expressing full-length recombinant P4H-TM likewise showed the presence of two immunoreactive polypeptides (Figure 3A in I). According to gel filtration analysis P4H-TM was eluted in fractions corresponding to molecular masses of 105-120 kDa and 85-90 kDa. This data suggests that both the full-length P4H-TM and its cleaved form exist as homodimers.

Purification of membrane proteins is troublesome, and therefore two additional baculovirus constructs were generated in order to obtain purified P4H-TM. These constructs lacked the transmembrane domain and either a GP67 or a PDI signal peptide followed by a His-tag was added to the N-terminus of P4H-TM. Both polypeptides were purified to homogeneity and used as a source of enzyme in activity assays.

As the C-P4Hs function within the lumen of the ER, while the HIF-P4Hs are cytoplasmic or nuclear enzymes, it was of interest to study the subcellular localization of the catalytic region of P4H-TM. Immuno EM studies were carried out with Sf9 and HEK293 cells expressing the full-length recombinant P4H-TM. The antibody used had been generated against a purified truncated P4H-TM beginning after the transmembrane domain at Asp88 and this thus revealed the location of the catalytic C-terminal region. P4H-TM was found to be an ER membrane protein located in an orientation in which its catalytic site was inside the lumen (Figure 5 in I). Endogenous P4H-TM in nontransfected HEK293 cells had the same location and orientation. This location and orientation was in agreement with the data indicating that P4H-TM was N-glycosylated (Figure 3B in I), and the presence of the -Arg-Val-Glu-Leu variant of the ER retention signal.

### **5.1.2 P4H-TM is expressed in many human tissues and cell lines and is induced by hypoxia**

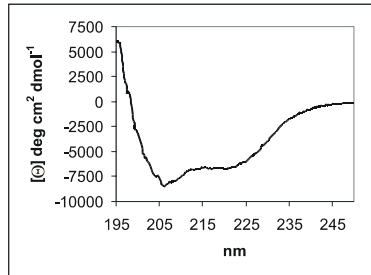
The mRNA expression pattern of P4H-TM revealed marked differences when compared to the C-P4Hs. The P4H-TM mRNA was expressed at high levels in the adult brain (Figure 2 in I), where the C-P4H  $\alpha$ (I)-(III) mRNAs are expressed only at low levels, if at all (Annunen *et al.* 1997, Kukkola *et al.* 2003). The high P4H-TM mRNA levels in the Northern blot analyses were also seen in the adult heart,

placenta, skeletal muscle, kidney and pancreas, with adult tissues showing higher levels than fetal tissues (Figure 2A-C in I). Three mRNA species of 2.7, 2.3 and 1.8 kb were seen, the 1.8-kb mRNA being the most abundant form in all tissues studied, except in the brain. P4H-TM mRNA was also expressed in epiphyseal cartilage and fibroblasts (Figure 2D in I) and in several tumour samples (Figure 2E in I).

Our Western blot studies showed that P4H-TM was expressed in a variety of human cell lines and its expression was markedly induced by hypoxia in most of them (Figure 4 in I). Analysis of the 5' untranslated sequence of the gene encoding human P4H-TM revealed the presence of five HREs thus supporting the evidence for hypoxia-inducibility. Immunofluorescence staining revealed that endogenous P4H-TM in normoxic and hypoxic HEK293 cells was located in the cytoplasm and showed a reticular staining pattern resembling the arrangement of the ER. In accordance with our other results, the immunofluorescence stainings also supported the proposed hypoxia-inducibility of the P4H-TM. However, hypoxia did not affect its cellular localization.

### **5.1.3 Actions of P4H-TM closely resemble those of HIF-P4Hs**

Based on circular dichroism analysis, the purified truncated P4H-TM was in a folded conformation (Fig. 5). Furthermore, like the C-P4Hs and HIF-P4Hs (Myllyharju & Kivirikko 1997, Hirsilä *et al.* 2003) it was found to catalyze uncoupled 2-oxoglutarate decarboxylation, although at a lower rate, indicating that P4H-TM was folded into a catalytically active conformation. We therefore set out to study the ability of P4H-TM to hydroxylate a wide range of potential peptide substrates using an assay based on the hydroxylation-coupled decarboxylation of 2-oxo[1-<sup>14</sup>C]glutarate (Kivirikko & Myllylä 1982). The non-purified recombinant full-length P4H-TM and the purified truncated P4H-TM beginning at Asp88 were used as sources of the enzyme. The peptides studied included collagen-like peptides such as (Pro-Pro-Gly)<sub>10</sub>, bradykinin analogues that are hydroxylated by the C-P4Hs *in vitro* (Kivirikko & Pihlajaniemi 1998), proline-rich peptides that act as *in vitro* substrates for viral and plant P4Hs (Eriksson *et al.* 1999, Hieta & Myllyharju 2002), and various 19-20 and 35-residue HIF- $\alpha$ -like peptides (Hirsilä *et al.* 2003, Koivunen *et al.* 2006). None of these served as a substrate for P4H-TM, however.



**Fig. 5. CD spectroscopy analysis of the purified His-tagged P4H-TM starting at Asp88 in the far UV region.**

The overexpression of P4H-TM has previously been shown to suppress cellular HIF-1 $\alpha$  and HIF-2 $\alpha$  reporter transactivation activity and to reduce the HIF-2 $\alpha$  protein level (Oehme *et al.* 2002). We therefore set out to study whether P4H-TM is involved in the hydroxylation of HIF-1 $\alpha$  and HIF-2 $\alpha$  ODDD*s in cellulo*. To do this we transiently transfected human neuroblastoma cells with a constant amount of a reporter plasmid for the N- or C-terminal half of the HIF-1 $\alpha$  or HIF-2 $\alpha$  ODDDs, together with increasing amounts of a plasmid coding for a full-length P4H-TM or HIF-P4H-2. Overexpression of both enzymes reduced the amounts of all four ODDD polypeptides in a dose-dependent manner and with about equal efficiency (Figure 6A in I). In contrast, silencing the genes for P4H-TM or any of the three HIF-P4Hs with siRNA oligonucleotides led to increased levels of HIF-1 $\alpha$  (Figure 6B in I). These data indicate that overexpression and silencing of P4H-TM leads to similar effects on the stability of HIF-1 $\alpha$  as overexpression and silencing of the HIF-P4Hs.

Oehme and coworkers (2002) were not able to show that recombinant P4H-TM is an active enzyme. Interestingly, it has recently been demonstrated that HIF-P4Hs require long substrates (Hirsilä *et al.* 2003) and their  $K_m$  values for recombinant HIF-1 $\alpha$  and HIF-2 $\alpha$  ODDDs consisting of more than 200 residues are lower than for 35-residue synthetic peptides (Koivunen *et al.* 2006). We decided to investigate the possibility that the synthetic peptides tested as substrates were not hydroxylated because of their short length. We studied hydroxylation of the 248-residue wild-type and mutant HIF-1 $\alpha$  ODDDs, the full-length HIF-1 $\alpha$  and the full-length pro $\alpha$ 1 chains of type I procollagen, which were produced in rabbit reticulocyte lysates as L-[2,3,4,5- $^3$ H]proline-labelled polypeptides by the purified truncated P4H-TM. The full-length HIF-1 $\alpha$  polypeptide and its ODDD were found to serve as substrates (Table 1 in I), while

the procollagen chains were not hydroxylated. The amount of 4-hydroxy[<sup>3</sup>H]proline formed was about 20% and 50% of those obtained, with an equal amount of HIF-P4H-2 and HIF-P4H-3, respectively (Table 1 in I). A Pro402Ala mutant HIF-1 $\alpha$  ODDD, was almost as good a substrate as the wild type, whereas hydroxylation of a Pro564Ala mutant was only about 10% of the wild type (Table 1 in I). P4H-TM thus resembles HIF-P4H-3 in that it preferentially hydroxylates the C-terminal Pro564 of HIF-1 $\alpha$  (Table I in I). A low but distinct level of 4-hydroxy[<sup>3</sup>H]proline formation was obtained with purified P4H-TM even when a Pro402,564Ala double mutant HIF-1 $\alpha$  ODDD was used as a substrate (Table 1 in I). This suggests that P4H-TM also acted at a low rate on some additional proline(s) that were not hydroxylated by the three HIF-P4Hs.

## **5.2 Molecular cloning and characterization of the second *Arabidopsis thaliana* prolyl 4-hydroxylase, At-P4H-2 (II)**

### **5.2.1 Cloning of At-P4H-2**

A sequence homology search of the *A. thaliana* genome has identified at least six polypeptides which show 21-27% identity to the catalytic C-terminal regions of the human C-P4H-I and II  $\alpha$  subunits (Hieta & Myllyharju 2002). One of the uncharacterized *A. thaliana* P4H-like polypeptides, named At-P4H-2 here, shared sequence homology with the human P4H-TM which was cloned and characterized in this study (original article I) within the potential peptide substrate binding region. The sequence identity of At-P4H-2 with the peptide substrate binding domain of the human C-P4H  $\alpha$ (I) subunit and the corresponding region of P4H-TM was 5% and 37%, respectively (Fig. 6). The P4H-TM likewise had only a 5% sequence identity with the peptide binding region of the C-P4H  $\alpha$ (I) subunit (Fig. 6). None of the tyrosine residues, which have been shown to be important for peptide binding in the case of the C-P4H  $\alpha$ (I) subunit (Hieta *et al.* 2003, Pekkala *et al.* 2004), were conserved in At-P4H-2 or P4H-TM. We regarded it possible that At-P4H-2 could turn out to be a beneficial tool for screening potential new substrate candidates for the human P4H-TM because it did not contain a transmembrane domain and like other plant P4Hs it was expected to be a soluble monomer. We therefore amplified the full-length At-P4H-2 cDNA from an *A. thaliana* Uni-ZAP XR cDNA library (Stratagene) by using primers designed according to the GenBank<sup>TM</sup> sequence NP566279.1.

```

C-P4H  $\alpha$  (I)      144  FLTAEDCFELGK[VAYTEADYYHTELWMEQALRQ]DEGEISTIDKVSVLDYLSYA
P4H-TM           135  -----IRTL[SLKPLLEIIPGFLTDEECRLI]-----TH
At-P4H-2         38  -----VKQV[S]SKPRA[V]VYEGFLTDL[EC]DH-----TS

VY[Q]GD[LD]KALL[LT]TKLLE[LD]PEHQ[R]ANGNLKYFEYIMAKEKDVNKS
LAQMKGLQRSQI[LP]TEEY[E]EAMSTMQVSQL-----
LAKEN-L[QRS]AVADNDNGE[SQV]SDV[R]TSSG-----

```

**Fig. 6. Sequence alignment of the peptide substrate binding domain of the human C-P4H  $\alpha$ (I) subunit with the corresponding regions of the P4H-TM and At-P4H-2 polypeptides. Identical residues are indicated with a black background. Gaps were introduced for maximal alignment.**

### **5.2.2 Expression of recombinant At-P4H-2 in insect cells**

The cDNA encoding At-P4H-2 was cloned into a baculovirus vector and used to generate a recombinant virus. *Spodoptera frugiperda* Sf9 insect cells were infected with the amplified virus and harvested 72 h after infection. The cells were homogenized in a buffer containing 0.1% Triton X-100, and centrifuged. The remaining pellet was solubilized consecutively in a buffer containing 1% Triton X-100, a buffer containing 50% glycerol, 0.6 M NaCl, and 1% Nonident P-40, and finally in 1% SDS. Aliquots of all samples were analyzed by SDS-PAGE. The 35 kDa At-P4H-2 polypeptide was poorly soluble in the Triton X-100 containing buffers whereas a strong band was seen in the glycerol and SDS extracts (Figure 2 in II). N-terminal sequencing showed that the At-P4H-2 polypeptide was present in all fractions, however.

### **5.2.3 Substrate specificity and catalytic properties of recombinant At-P4H-2**

P4H activity was assayed using a method based on the hydroxylation-coupled decarboxylation of 2-oxo[1-<sup>14</sup>C]glutarate (Kivirikko & Myllylä 1982). A significant amount of P4H activity was observed in the 0.1% Triton X-100-soluble fraction of cells expressing At-P4H-2 when poly(L-proline) was used as the peptide substrate (Table I in II). The observed At-P4H-2 activity represented hydroxylation-coupled decarboxylation and not uncoupled decarboxylation as shown by analysis of the amount of 4-hydroxyproline formed (Table II in II). Although the amount of recombinant At-P4H-2 was much higher in the glycerol-soluble fraction than in the 0.1% Triton X-100-soluble fraction, the P4H activity

in it was only 16% of that in the Triton X-100 fraction. This suggests that the majority of the glycerol-soluble At-P4H-2 was incorrectly folded. The 0.1% Triton X-100-soluble fraction was therefore used as the source of At-P4H-2 in all enzymatic analyses.

A wide range of synthetic peptides were tested as potential substrates for At-P4H-2 (Table I in II). A large amount of P4H activity was obtained with poly(L-proline) of various lengths, a peptide representing an AGP sequence and two peptides representing sequences found in extensins (Table I in II). Hydroxylation of poly(L-proline) was more effective, however, as its  $K_m$  values were markedly lower than those of the AGP and extensin peptides (Table III in II). The  $K_m$  values of At-P4H-2 for poly(L-proline),  $M_r$  30,000-40,000 and  $M_r$  5,000-10,000, were 13 and 30  $\mu\text{M}$ , respectively, being thus much higher than those of the previously cloned *A. thaliana* P4H, At-P4H-1, the  $K_m$  of which for poly(L-proline),  $M_r$  10,000-20,000, was as low as 0.2  $\mu\text{M}$  (Table III in II). The  $K_m$  values of At-P4H-2 for the AGP and extensin peptides ranged from 380 to 1260  $\mu\text{M}$ , being 10-60 times higher than those of At-P4H-1 (Table III in II). N-terminal sequencing of the 21-residue extensin-like peptide SPPPVYKSPPPPVKHYSPPPV after hydroxylation by At-P4H-2 showed that the first proline of the central proline quadruplet and the most C-terminal proline were hydroxylated, the hydroxylation level of the C-terminal proline being higher (Figure 3 in II).

Although plants do not have collagens, the collagen-like peptide (Pro-Pro-Gly)<sub>10</sub> has been shown to be a good substrate for At-P4H-1 (Hieta & Myllyharju 2002). In contrast, At-P4H-2 hydroxylated (Pro-Pro-Gly)<sub>10</sub> very inefficiently, thus resembling the viral PBCV-1 P4H in this respect (Tables I and III in II). The single -Pro-Pro-Gly- triplet in bradykinin, a peptide involved in the regulation of blood pressure, was not hydroxylated by At-P4H-2 (Table I in II). The prion protein has been shown to contain one 4-hydroxyproline residue and its N terminus is predicted to adopt a poly(L-proline) type II helix conformation (Gill *et al.* 2000). A low level of At-P4H-2 activity was generated with a peptide representing the hydroxylation site of the prion protein (Table I in II). At-P4H-1 has been shown to hydroxylate synthetic peptides representing the hydroxylation sites of HIF-1 $\alpha$  (Hieta & Myllyharju 2002), but At-P4H-2 did not hydroxylate any of the HIF- $\alpha$  peptides tested (Table I in II).

The At-P4H-2 required  $\text{Fe}^{2+}$ , 2-oxoglutarate,  $\text{O}_2$  and ascorbate for activity, as expected. Its  $K_m$  values for the cosubstrates were similar to those of At-P4H-1 (Table IV in II). The  $K_m$  value for  $\text{Fe}^{2+}$  was 5  $\mu\text{M}$ , which was about one-third of that of At-P4H-1 but about 2 and 10-fold higher than those of the human C-P4H-I,

and the PBCV-1 P4H, respectively (Table IV in II). The  $K_m$  value for 2-oxoglutarate, 170  $\mu\text{M}$ , was slightly higher than that of At-P4H-1, both of these values being much higher than those of C-P4H-I and PBCV-1 P4H (Table IV in II). The  $K_m$  values for ascorbate were identical for At-P4Hs, C-P4H-I and PBCV-1 P4H (Table IV in II).

Two well known competitive inhibitors of the vertebrate C-P4Hs that compete with 2-oxoglutarate – pyridine 2,4-dicarboxylate and pyridine 2,5-dicarboxylate – also acted as inhibitors for At-P4H-2, although the  $K_i$  values were higher than those of C-P4H-I and PBCV-1 P4H (Table IV in II). Pyridine 2,4-dicarboxylate was a more efficient inhibitor of At-P4H-2 than pyridine 2,5-dicarboxylate, and the relative inhibition efficiency of these compounds is the opposite with C-P4H-I and PBCV-1 P4H (Table IV in II).

### **5.3 Characterization of the human prolyl 3-hydroxylase isoenzyme 2 (III)**

#### **5.3.1 Cloning and expression of the recombinant P3H isoenzymes in insect cells**

In order to express the three human P3H isoenzymes as recombinant proteins, PCR amplification was used to obtain the corresponding full-length cDNAs from different human cDNA pools with specific primers. The P3H isoenzymes were expressed as recombinant proteins in insect cells. The cells were harvested 72 h after infection and homogenized, and the Triton X-100 and SDS-soluble fractions were analyzed by SDS-PAGE followed by Coomassie blue staining or Western blotting (Figure 2 in III) with polyclonal antibodies generated against synthetic mouse P3H1 and P3H2 peptides. The recombinant P3H1 and P3H2 polypeptides were present in the Triton X-100-soluble fraction, although the majority of them were soluble only in 1% SDS (Figure 2 in III). P3H3 seemed to be completely insoluble in the Triton X-100 buffer (Figure 2A in III). The presence of the P3H1 and P3H2 polypeptides in the Triton X-100-soluble fraction was verified by N-terminal sequencing. Furthermore, N-terminal sequencing showed that the signal peptides of P3H1, 2 and 3 consist of 18, 24 and 20 amino acids, respectively.

### 5.3.2 Substrate specificity and catalytic properties of recombinant P3H2

The commonly used activity assay for P4Hs (Myllylä & Kivirikko 1982, Hirsilä *et al.* 2003), based on the hydroxylation-coupled decarboxylation of 2-oxo-[<sup>14</sup>C]glutarate, was adopted here to study the activity and catalytic properties of the recombinant P3Hs. It has been previously shown that prolyl 3-hydroxylation most likely requires the presence of 4-hydroxyproline in the Y position, i.e. a –Pro-4Hyp-Gly- sequence (Tryggvason *et al.* 1977, Risteli *et al.* 1977). Therefore, a 15-amino-acid peptide (Pro-4Hyp-Gly)<sub>5</sub> was used here as a substrate to measure P3H activity. Using this assay, P3H activity was obtained only in the Triton X-100-soluble fractions of cells expressing P3H2 (Table 1 in III). The lack of enzyme activity in the case of P3H1 was most probably due to several point mutations that were found in the P3H1 sequence when it was compared to sequences available in databases. Several versions of the P3H1 cDNA were produced by correcting the point mutations in different combinations by site-directed mutagenesis. However, none of the recombinant P3H1 polypeptides had enzyme activity. The lack of enzyme activity in the case of P3H3 was most likely a consequence of its extremely poor solubility in the Triton X-100 buffer. As it has been recently published that the cartilage associated protein, CRTAP, is associated with P3H1 and is required for efficient prolyl 3-hydroxylation *in vivo* (Vranka *et al.* 2004, Barnes *et al.* 2006, Morello *et al.* 2006), we cloned CRTAP and coexpressed it with all P3H isoenzymes. Coexpression with CRTAP did not increase the solubility or activity of any of the three isoenzymes.

The kinetic constants of P3H2 for the peptide substrate, the reaction cosubstrates and certain inhibitors were determined. P3H2 efficiently hydroxylated the (Gly-Pro-4Hyp)<sub>5</sub> substrate, the  $K_m$  value of 70  $\mu$ M being lower than that of C-P4H-I for a peptide substrate of a similar length (Table 2 in III). The  $K_m$  values of P3H2 for the cosubstrates also had certain differences when compared to those of the human C-P4Hs. The  $K_m$  of P3H2 for Fe<sup>2+</sup> was 0.5  $\mu$ M, which is four times lower than that of C-P4H-I (Table 2 in III). In contrast, the  $K_m$  for 2-oxoglutarate, 80  $\mu$ M, was four times higher than that of C-P4H-I (Table 2 in III). The  $K_m$  of P3H2 for ascorbate was about one third of that of C-P4H-I (Table 2 in III).

Like C-P4Hs, P3H2 was efficiently inhibited by pyridine 2,4-dicarboxylate with a  $K_i$  value of 9  $\mu$ M (Table 2 in III). In contrast, pyridine 2,5-dicarboxylate which is an even more efficient inhibitor of the C-P4Hs than pyridine 2,4-

dicarboxylate, was a very poor inhibitor of P3H2 (Table 2 in III). Poly(L-proline) also inhibited P3H2 very inefficiently (Table 2 in III).

### **5.3.3 Expression of the P3H isoenzymes in vertebrate tissues**

The P3H1 mRNA was expressed in all adult and fetal tissues studied, the highest expression levels being in placenta, adult lung, liver, kidney and pancreas, and in fetal spleen, lung, liver, skeletal muscle and kidney (Figure 1 in III). The P3H2 mRNA was likewise expressed in all tissues studied except for adult brain and skeletal muscle, with the highest expression levels in placenta, adult lung, liver and kidney, and fetal heart, spleen, lung, liver, skeletal muscle and kidney (Figure 1 in III). The P3H3 mRNA expression closely resembled that of P3H1 in adult tissues, but was more restricted in the fetal tissues, where the highest expression was seen in the lung, skeletal muscle and kidney (Figure 1 in III).

The antibodies produced against synthetic peptides representing the mouse P3H1 and P3H2 polypeptides were used to study the expression of these isoenzymes in adult mouse tissues. The antibodies were found to be highly isoenzyme-specific, as shown in the Western blot analysis of the recombinant P3H1 and P3H2 polypeptides (Figure 2B and 2C in III). It has been shown that 3-hydroxyproline is most abundantly found in the type IV collagen, which is one of the major components in basement membranes (Gryder *et al.* 1975, Kefalides 1973, Kresina & Miller 1979) but the chick P3H1 has previously been shown to be expressed only in tissues where fibrillar collagens predominate (Vranka *et al.* 2004). Thus it was of interest to study whether P3H2 would be expressed in tissues rich in basement membranes.

In general, P3H2 was found to be expressed strongly in tissues where basement membranes and thus collagen IV predominate, while P3H1 was expressed only at a low level, if at all (Figure 3 in III). P3H2 was strongly expressed in the tubular cells of the kidney (Figure 3 in III). Immuno EM revealed that P3H2 was present in the kidney tubular cells close to the basement membrane in an apparent association with membranous structures, particularly in the basal folds of the tubular cells (Figure 4 in III). In agreement with the previous studies (Vranka *et al.* 2004), P3H1 showed no staining in the kidney (Figure 3 in III). P3H2 was also strongly expressed in the Schwann cells of the peripheral nerve whereas there was no P3H1 staining seen. Both P3H1 and P3H2 were expressed in the pancreas but had dissimilar staining patterns (Figure 3 in III). P3H2 was expressed in the acinar cells and in the cuboidal epithelium cells

of the interlobular ducts. In contrast, P3H1 was expressed in the walls of intercalated ducts, which consist of a flattened cuboidal epithelium. Both P3H1 and P3H2 were expressed in the tunica adventitia, the smooth muscle layer of the aortic wall (Figure 3 in III). P3H1 has previously been shown to be expressed in the tunica adventitia as well (Vranka *et al.* 2004). In the liver, P3H1 appeared to be expressed in the canaliculi formed by adjacent hepatocytes (Figure 3 in III). P3H2 was not expressed in the liver. No staining with either of the antibodies was seen in the heart, skeletal muscle or brain.



## 6 Discussion

### 6.1 The actions of P4H-TM resemble those of HIF-P4Hs but it may also act on substrates other than HIF-1 $\alpha$

P4H-TM was identified from the GeneBank database on the basis of its sequence homology to the catalytic subunits of C-P4Hs. Although the sequence identity of the C-terminal catalytic region of P4H-TM was higher with the C-P4H  $\alpha$  subunit than the HIF-P4H sequences, and although its catalytic sites were located within the lumen of the ER, P4H-TM failed to hydroxylate recombinant procollagen polypeptide chains, whereas in the same experiments it hydroxylated HIF-1 $\alpha$  and its ODDD. This failure to hydroxylate procollagen chains may be explained by the fact that P4H-TM lacks any sequence similarity with the peptide substrate binding domain of the C-P4Hs  $\alpha$  subunits (Myllyharju & Kivirikko 1999). Furthermore, the mRNA expression pattern of P4H-TM was distinct from those of the C-P4H  $\alpha$  subunits, the most remarkable difference being the high P4H-TM mRNA level in the brain where, in contrast, the expression of C-P4H  $\alpha$  subunits is very low.

Overall, the actions of P4H-TM closely resembled those of HIF-P4Hs. Our studies in several mammalian cell lines suggested that it was hypoxia-inducible in most of them thus resembling HIF-P4H-2 and 3 (Berra *et al.* 2003, Metzen *et al.* 2003a). The HRE sequences found in the 5' untranslated region of the P4H-TM gene also supported the idea that P4H-TM is hypoxia-inducible. P4H-TM was shown to hydroxylate the two critical prolines in a recombinant [<sup>3</sup>H]proline-labelled HIF-1 $\alpha$  and its ODDD *in vitro*, although P4H-TM does not share marked homology with the 17-residue sequences of HIF-P4H-2 and HIF-P4H-3 that have been shown to be necessary for HIF-1 $\alpha$  ODDD binding (Villar *et al.* 2007). In addition to the prolines hydroxylated by HIF-P4Hs, 21 proline residues are found in the HIF-1 $\alpha$  ODDD. Interestingly, P4H-TM also acted at a low rate on some of these other proline residues. It has been speculated that HIF-P4Hs may have additional physiological substrates, besides the HIF- $\alpha$ s (Huang LE *et al.* 2002, Metzen & Ratcliffe 2004, Dann & Bruick 2005), as their *in vitro* substrate requirements are not very strict (Huang LE *et al.* 2002, Li *et al.* 2004). However, to date there is only indirect evidence for such new substrate candidates, e.g. iron regulatory protein-2, the large subunit of RNA polymerase II, I $\kappa$ B kinase- $\beta$  and activating transcription factor 4 (Kuznetsova *et al.* 2003, Wang & Pantopoulos

2005, Cummins *et al.* 2006, Köditz *et al.* 2007). In addition, the recent report that HIF-P4H-3 has a novel role in promoting apoptosis during development (Lee *et al.* 2005) suggests that other functions of HIF-P4Hs remain to be found. Interestingly, FIH has also been shown to have other functions, in addition to the regulation of HIF activity, as it hydroxylates ankyrin repeats in I $\kappa$ B proteins and the Notch receptors (Cockmann *et al.* 2006, Coleman *et al.* 2007). In these studies it was proposed that FIH-mediated HIF- $\alpha$  hydroxylation is competitively inhibited through the hydroxylation of asparagine residues of the ankyrin repeat containing proteins, thus revealing another oxygen-dependent regulation system for HIF signalling.

Overexpression of P4H-TM was shown to reduce the recombinant HIF- $\alpha$  ODDD levels by the same extent as overexpression of HIF-P4Hs and silencing of endogenous P4H-TM by siRNA increased the HIF-1 $\alpha$  level in a manner very similar to that achieved by silencing of HIF-P4Hs. However, the subcellular localization of P4H-TM was found to be distinctly different from that of the HIF-P4Hs as our immuno EM data clearly demonstrated that P4H-TM was localized in the ER membranes in an orientation in which its catalytic site was inside the lumen. This result was also supported by the findings that P4H-TM was N-glycosylated and that its C terminus had a functional variant of the ER retention signal. In contrast, HIF-P4Hs are localized in the nucleus and cytoplasm, suggesting that HIF- $\alpha$  is hydroxylated in those cellular compartments (Huang J *et al.* 2002, Metzén *et al.* 2003a). The levels of HIF- $\alpha$ s in normoxic cells are so low that they are scarcely detectable but a number of studies have demonstrated that in hypoxia HIF- $\alpha$  is located in the nucleus (Jiang *et al.* 1996, Kallio *et al.* 1998). This creates a dilemma: How was the transfected ER-localized P4H-TM able to act on HIF- $\alpha$  ODDDs in the same way as the transfected HIF-P4H-2? And how did the silencing of the endogenous P4H-TM by siRNA influence the HIF-1 $\alpha$  protein level in a similar manner to silencing of HIF-P4Hs? However, a few studies are available which show that a substantial amount of HIF- $\alpha$ s are also found in the ER in normoxia (Liu *et al.* 2004). Interestingly, there is also data available indicating that HIF-1 $\alpha$  and HIF-P4Hs are targeted to the peroxisomes in hypoxic primary hepatocytes, suggesting that some of these polypeptides also exist in membranous cytoplasmic organelles (Khan *et al.* 2006). In addition, it is known that the outer nuclear membrane is continuous with the membranes of the ER (Aitchison & Rout 2002), which could perhaps also explain the actions of P4H-TM on HIF- $\alpha$ s. In the light of our study and the studies discussed above, it is possible that P4H-TM may have additional physiological substrates and that it

may play an even more important role in the hydroxylation of those substrates than in the regulation of HIF- $\alpha$ s *in vivo*.

## 6.2 At-P4H-2 and At-P4H1 have distinct substrate specificities

The recombinant At-P4H-2 was found to be a monomer of about 35 kDa, which showed significant P4H activity when poly(L-proline) was used as a substrate. The other previously characterized plant P4Hs and viral PBCV-1 P4H have also been found to exist as monomers and have been shown to hydroxylate poly(L-proline) efficiently, thus differing markedly from vertebrate C-P4Hs (Kaska *et al.* 1987, 1988; Eriksson *et al.* 1999, Hieta & Myllyharju 2002, Keskiäho *et al.* 2007).

Altogether, four plant P4Hs have now been cloned and characterized, At-P4H-2 being the second one, preceded by the cloning of At-P4H-1 (Hieta & Myllyharju 2002) and followed by the *N. tabacum* and *C. reinhardtii* P4Hs (Yuasa *et al.* 2005, Keskiäho *et al.* 2007). The amino acid sequence identity between At-P4H-2 and At-P4H-1 was 33%, the most obvious difference being a 49-residue C-terminal extension, a toxin-like domain with 6 cysteines, in At-P4H-2. This extension was also found in a putative rice homologue of At-P4H-2 that shared 51% identity with At-P4H-2, and interestingly also in the splicing variant 1B of the *C. reinhardtii* P4H-1 (Keskiäho *et al.* 2007). A truncated Cr-P4H-1, which lacked the C-terminal toxin-like domain was a fully active P4H, however (Keskiäho *et al.* 2007).

At-P4H-2 was found to differ greatly from At-P4H-1 in its substrate specificity. At-P4H-2 hydroxylated poly(L-proline) and the plant HRGP peptides much less efficiently than At-P4H-1. Furthermore, unlike At-P4H-1, At-P4H-2 did not hydroxylate HIF- $\alpha$ -like peptides at all and it hydroxylated a collagen-like peptide only very inefficiently. The N-terminal region of At-P4H-2 contained a stretch of 24 amino acids in which its sequence identity to the rice homologue and the viral PBCV-1 P4H was much higher than to At-P4H-1. This area could present a potential peptide binding region and might explain at least some of the differences found between the substrate specificities of At-P4H-2 and PBCV-1 P4H and that of At-P4H-1. However, it is still unclear whether the small monomeric plant P4Hs have a distinct substrate binding domain or whether the substrates become bound straight to their catalytic domain.

The synthetic peptides that were efficiently hydroxylated by At-P4H-2 contained at least three consecutive prolines and N-terminal sequencing of the partially hydroxylated extensin-like peptide SPPPVYKSPPPPVKHYSPPPV

showed that At-P4H-2 acted on the 1<sup>st</sup> proline of the central quadruplet and the 3<sup>rd</sup> proline of the C-terminal triplet, the hydroxylation level of the latter being higher. It thus seems that the minimum substrate requirement of At-P4H-2 is fulfilled by three consecutive prolines. At-P4H-2 was originally cloned and analyzed in this study with the idea that it might shed light on the substrate requirements of the human P4H-TM. This idea arose because At-P4H-2 and human P4H-TM have sequence similarity in the region corresponding to the peptide substrate binding domain of the C-P4H  $\alpha$  subunits. At this point, it must be concluded that the substrate specificities of At-P4H-2 and P4H-TM are distinct and At-P4H-2 can not be used as a model to study P4H-TM. Instead, analysis of At-P4H-2 showed for the first time that individual members of a large plant P4H family have specific substrate requirements.

As At-P4H-2 hydroxylated various proline-rich peptides representing plant HRGP sequences much less efficiently than At-P4H-1, it remains to be discovered whether At-P4H-2 has any major role in their hydroxylation *in vivo*. Recent data suggests that individual members of plant P4H families are likely to have specific *in vivo* functions, as a lack of Cr-P4H-1 activity could not be replaced by the nine other *C. reinhardtii* P4H-like polypeptides and led to abnormal cell wall assembly (Keskiäho *et al.* 2007).

The  $K_m$  values of At-P4H-2 for cosubstrates were highly similar to those of the other enzymatically characterized plant P4Hs (Hieta & Myllyharju 2002, Keskiäho *et al.* 2007). The most notable difference in cosubstrate requirements between plant and vertebrate P4Hs is the relatively high  $K_m$  values of the plant P4Hs for iron and 2-oxoglutarate, as presented here and in other studies (Tables 1-3 and Hieta & Myllyharju 2002, Keskiäho *et al.* 2007). The differences in the  $K_m$  values for 2-oxoglutarate and  $K_i$  values for 2-oxoglutarate analogue inhibitors between the plant P4Hs, C-P4Hs and HIF-P4Hs (Tables 1-3 and original article II) suggest that there are some distinct differences in the structures of the 2-oxoglutarate binding sites of these enzymes.

### **6.3 P3H2 is associated with the hydroxylation of basement membrane collagens**

The three human P3H isoenzymes were cloned and produced as recombinant proteins in order to study their enzyme kinetics. Only one of the isoenzymes, P3H2, was expressed as an active recombinant protein, however. The lack of enzyme activity in the other two isoforms was probably due to the number of

mutations found in the P3H1 cDNA and the poor solubility of P3H3. CRTAP has been shown to form a complex with P3H1 and it is required for efficient *in vivo* prolyl 3-hydroxylation of fibrillar collagens (Vranka *et al.* 2004, Barnes *et al.* 2006, Morello *et al.* 2006). Coexpression of CRTAP did not increase the solubility or activity of any of the recombinant P3H polypeptides, however. Our results thus suggest that CRTAP does not function as a chaperone for the P3H polypeptides, but its function may rather be to ensure correct interaction with collagen polypeptides.

One of the important aspects of this study was the development of a simple assay for the study of P3H activity. The principle of this assay is the same as that of the assays that are routinely used to study C-P4H and collagen lysyl hydroxylase (LH) activity, i.e. determination of the amount of the hydroxylation-coupled decarboxylation of radioactively labelled 2-oxoglutarate when a short synthetic peptide is used as a substrate, which in the case of P3Hs is a (Pro-4Hyp-Gly)<sub>5</sub> peptide. The recombinant P3H2 hydroxylated the (Pro-4Hyp-Gly)<sub>5</sub> peptide efficiently but did not act on the C-P4H peptide substrate (Pro-Pro-Gly)<sub>5</sub>. This finding confirmed the previous observations that prolyl 3-hydroxylation of X position prolines is dependent on the presence of 4-hydroxyproline in the Y position (Tryggvason *et al.* 1977, Risteli *et al.* 1977). It would be of interest to study the *in vivo* actions of P3Hs and C-P4Hs on collagen polypeptides, in particular, how they are regulated and how they interact.

Our results showed that there were some distinct differences between P3H2 and the C-P4Hs in their  $K_m$  values for cosubstrates. The most notable of these were in the  $K_m$  and  $K_i$  values for 2-oxoglutarate and 2-oxoglutarate analogue inhibitors, respectively. For example, pyridine 2,5-dicarboxylate, which is a very efficient inhibitor of C-P4Hs, was a poor inhibitor of P3H2. In these properties, P3H2 actually resembled LHs (for a review, see Myllyharju 2005) more closely than C-P4Hs. These differences are likely to be due to differences in the structures of the 2-oxoglutarate binding sites; P3H2 and LHs share certain common features that are distinct from those of C-P4Hs. One of these features is that the residue that binds the C5 carboxyl group of 2-oxoglutarate is an arginine in P3Hs and LHs, while it is a lysine in C-P4Hs (Myllyharju & Kivirikko 1997, Passoja *et al.* 1998, Vranka *et al.* 2004).

The most important finding of the present data was that the tissue expression pattern of P3H2 differed distinctly from that of P3H1. P3H2 was prominently expressed in tissues rich in basement membranes, e.g. in the kidney. This result was remarkable, as the amount of 3-hydroxyproline is much higher in basement

membrane type IV collagen than in fibrillar collagens (Kefalides 1973), and because P3H1 has been shown to be expressed only in tissues that express fibrillar collagens (Vranka *et al.* 2004). P3H1 has been shown to be responsible for the 3-hydroxylation of type I collagen *in vivo* as mutations in the human gene for P3H1 lead to a bone disorder resembling lethal or severe OI (Cabral *et al.* 2007). A lack of the single 3-hydroxyproline residue in collagen I polypeptides has been found to interfere with the proper assembly and secretion of type I collagen molecules (Barnes *et al.* 2006, Morello *et al.* 2006, Cabral *et al.* 2007). P3H2 was found to be localized in tissues rich in basement membranes, and this would suggest that this isoenzyme is responsible for the prolyl 3-hydroxylation of type IV collagen polypeptides. The malfunction of this isoenzyme could thus possibly lead to the diseases characterized by alterations in basement membranes.

## 7 Conclusions and future prospects

A novel human P4H, P4H-TM, was shown to act on HIF- $\alpha$ s in a similar manner to HIF-P4Hs suggesting that it is a fourth P4H that is involved in the oxygen-dependent regulation of the stability of HIF- $\alpha$ . Even though the *in vitro* and *in cellulo* actions of P4H-TM were similar to those of the HIF-P4Hs, the cellular localization of its catalytic site was the same as those of C-P4Hs and its sequence identity was higher with the C-P4H catalytic  $\alpha$  subunits than with HIF-P4Hs. Furthermore, it was capable of hydroxylating a mutant HIF-1 $\alpha$  ODDD where both known HIF-P4H target prolines were mutated to an alanine, although at a low rate. These data suggest that P4H-TM may have some other physiological substrates and that it may play an even more important function in the hydroxylation of those substrates than in the regulation of HIF- $\alpha$ s *in vivo*. The putative peptide binding region of P4H-TM shared a marked homology with At-P4H-2, which was the major reason for the characterization of this second *A. thaliana* P4H in this study. The hydroxylation specificities of P4H-TM and At-P4H-2 were found to be distinct, however, and therefore At-P4H-2 did not turn out to be a useful tool for identifying potential novel candidate substrates for P4H-TM. On the other hand, characterization of At-P4H-2 showed for the first time that different members of plant P4H families have distinct substrate specificities and thus may have specific *in vivo* functions. This was verified recently in the case of *C. reinhardtii* where the function of Cr-P4H-1 in proper cell wall assembly can not be replaced by the other nine members of its P4H family.

Sequence database searches showed that a P4H-TM-like polypeptide with a 51% identity to the human sequence is found in the zebrafish but not in *C. elegans* or *D. melanogaster*. This suggests that potential new substrate candidates will only be found in vertebrates. Future approaches for the identification of other potential P4H-TM substrates could involve, for example, yeast two-hybrid screens and immunoprecipitation studies from cells treated with a P4H-TM inhibitor, which could, potentially, prolong their half-life thus enriching the amount of the enzyme-substrate complex. One approach that has already been initiated is a mouse model of the P4H-TM gene inactivation. The mouse model is particularly interesting in the light of the unique tissue expression pattern of P4H-TM, especially the high level in the brain.

P3H2 was shown here to be expressed at high levels in basement membrane rich tissues. The fibrillar collagen polypeptides contain only a single 3-hydroxyproline, while the number is much higher in type IV collagen

polypeptides. It would therefore be very interesting to study the *in vivo* role of this modification in type IV collagen by generating and analyzing gene-modified mice for P3H2. It is also possible that human mutations in the gene for P3H2 could lead to a disease phenotype characterized by changes in basement membranes.

It has been suggested that C-P4Hs could be potential targets for the prevention of excess collagen accumulation in fibrotic diseases and thus they have been intensively studied. The discovery of HIF-P4Hs identified a new aspect of HIF regulation and HIF-P4Hs have also been identified as extremely attractive targets for controlling HIF activity for therapeutic purposes. Pharmacological inhibition of C-P4Hs and HIF-P4Hs could provide a potential therapeutic target for the treatment of conditions that occur as a consequence of excess collagen accumulation, and ischemia or anaemia, respectively. As C-P4Hs, HIF-P4Hs and P3Hs require the same cosubstrates for their function it is important to understand the specifics of the catalytic properties of all these prolyl hydroxylases to be able to produce specific inhibitors in the future. This is a demanding job and this thesis study provides important data for it, as the kinetic properties of a P3H isoenzyme were characterized in detail for the first time in this study. The results showed that it will likely be possible to develop specific inhibitors for each group of prolyl hydroxylases. However, the identification and characterization of the novel P4H-TM and the possibility that it may have some as yet unknown *in vivo* substrates brings new information and new challenges to the development of specific P4H inhibitors for medical purposes.

## References

- Abrams EW & Andrew DJ (2002) Prolyl 4-hydroxylase  $\alpha$ -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 alivary gland lumen through regulation of two downstream target genes, PH4 $\alpha$ SG1 and PH4 $\alpha$ SG2. *Development* 133: 3517-3527.
- 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.
- Aitchison JD & Rout MP (2002) A tense time for the nuclear envelope. *Cell* 108: 301-304.
- 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  $\alpha$  subunit isoform  $\alpha$ (II) and characterization of the type II enzyme tetramer. The  $\alpha$ (I) and  $\alpha$ (II) subunits do not form a mixed  $\alpha$ (I) $\alpha$ (II) $\beta_2$  tetramer. *J Biol Chem* 272: 17342-17348.
- Annunen P, Koivunen P & Kivirikko KI (1999) Cloning of the  $\alpha$  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.
- Appelhoff RJ, Tian YM, Raval RR, Turley H, Harris AL, Pugh CW, Ratcliffe PJ & Gleadle JM (2004) Differential function of the prolyl hydroxylases PHD1, PHD2, and PHD3 in the regulation of hypoxia-inducible factor. *J Biol Chem*. 279: 38458-38465.
- Aprelikova O, Chandramouli GV, Wood M, Vasselli JR, Riss J, Maranchie JK, Linehan WM & Barrett JC (2004) Regulation of HIF prolyl hydroxylases by hypoxia-inducible factors. *J Cell Biochem* 92: 491-501.
- Aravind L & Koonin EV (2001) The DNA-repair protein AlkB, EGL-9, and Iprecan define new families of 2-oxoglutarate- and iron-dependent dioxygenases. *Genome Biol* 2: research0007.1-0007.8.
- Barnes AM, Chang W, Morello R, Cabral WA, Weis M, Eyre DR, Leikin S, Makareeva E, Kuznetsova N, Uveges TE, Ashok A, Flor AW, Mulvihill JJ, Wilson PL, Sundaram UT, Lee B & Marini JC (2006) Deficiency of cartilage-associated protein in recessive lethal osteogenesis imperfecta. *N Engl J Med* 355: 2757-2764.
- Bassuk JA, Kao WW, Herzer P, Kedersha NL, Seyer J, DeMartino JA, Daugherty BL, Mark GE 3<sup>rd</sup> & Berg RA (1989) Prolyl 4-hydroxylase: molecular cloning and the primary structure of the  $\alpha$  subunit from chicken embryo. *Proc Natl Acad Sci USA* 86: 7382-7386.
- Bella J, Brodsky B & Berman HM (1995) Hydration structure of a collagen peptide. *Structure* 3: 893-906.

- 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.
- 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.
- Berra E, Benizri E, Ginouvès A, Volmat V, Roux D & Pouyssegur J (2003) HIF prolyl-hydroxylase 2 is the key oxygen sensor setting low steady-state levels of HIF-1 $\alpha$  in normoxia. *EMBO J* 22: 4082-4090.
- 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.
- Bos KJ, Rucklidge GJ, Dunbar B & Robins SP (1999) Primary structure of the helical domain of porcine collagen X. *Matrix Biol* 18: 149-153.
- Bozas SE, Panaccio M, Creaney J, Dosen M, Parsons JC, Vlasuk GV, Walker ID & Spithill TW (1995) Characterisation of a novel Kunitz-type molecule from the trematode *Fasciola hepatica*. *Mol Biochem Parasitol* 74: 19-29.
- Bozas SE & Spithill TW (1996) Identification of 3-hydroxyproline residues in several proteins of *Fasciola hepatica*. *Exp Parasitol* 82: 69-72.
- Brahimi-Horn MC & Poysségur J (2007) Harnessing the hypoxia-inducible factor in cancer and ischemic disease. *Biochem Pharmacol* 73: 450-457.
- Bruegge K, Jelkmann W & Metzner E (2007) Hydroxylation of hypoxia-inducible transcription factors and chemical compounds targeting the HIF- $\alpha$  hydroxylases. *Curr Med Chem* 14: 1853-1862.
- Bruick RK & McKnight SL (2001) A conserved family of prolyl-4-hydroxylases that modify HIF. *Science* 294: 1337-1340.
- Burjanadze TV (1979) Hydroxyproline content and location in relation to collagen thermal stability. *Biopolymers* 18: 931-938.
- Cabral WA, Chang W, Barnes AM, Weis M, Scott MA, Leikin S, Makareeva E, Kuznetsova NV, Rosenbaum KN, Tiffit CJ, Bulas DI, Kozma C, Smith PA, Eyre DR & Marini JC (2007) Prolyl 3-hydroxylase 1 deficiency causes a recessive metabolic bone disorder resembling lethal/severe osteogenesis imperfecta. *Nat Genet* 39: 359-365.
- Cassab GI (1998) Plant cell wall proteins. *Ann Rev Plant Phys Plant Mol Biol* 49: 281-309.
- Cervera AM, Apostolova N, Luna-Crespo F, Sanjuan-Pla A, Garcia-Bou R & McCreath KJ (2006) An alternatively spliced transcript of the PHD3 gene retains prolyl hydroxylase activity. *Cancer Lett* 233: 131-138.
- Chan DA, Sutphin PD, Yen SE & Giaccia AJ (2005) Coordinate regulation of the oxygen-dependent degradation domains of hypoxia-inducible factor 1 $\alpha$ . *Mol Cell Biol* 25: 6415-6426.
- Cioffi CL, Liu XQ, Kosinski PA, Garay M & Bowen BR (2003) Differential regulation of HIF-1 $\alpha$  prolyl-4-hydroxylase genes by hypoxia in human cardiovascular cells. *Biochem Biophys Res Commun* 303: 947-953.

- Clifton IJ, McDonough MA, Ehrismann D, Kershaw NJ, Granatino N & Schofield CJ (2006) Structural studies on 2-oxoglutarate oxygenases and related double-stranded  $\beta$ -helix fold proteins. *J Inorg Biochem* 100: 644-669.
- Cockman ME, Lancaster DE, Stolze IP, Hewitson KS, McDonough MA, Coleman ML, Coles CH, Yu X, Hay RT, Ley SC, Pugh CW, Oldham NJ, Masson N, Schofield CJ & Ratcliffe PJ (2006) Posttranslational hydroxylation of ankyrin repeats in I $\kappa$ B proteins by the hypoxia-inducible factor (HIF) asparaginyl hydroxylase, factor inhibiting HIF (FIH). *Proc Natl Acad Sci USA* 103: 14767-14772.
- Coleman ML, McDonough MA, Hewitson KS, Coles C, Mecinovic J, Edelmann M, Cook KM, Cockman ME, Lancaster DE, Kessler BM, Oldham NJ, Ratcliffe PJ & Schofield CJ (2007) Asparaginyl hydroxylation of the Notch ankyrin repeat domain by factor inhibiting hypoxia-inducible factor. *J Biol Chem* 282: 24027-24038.
- Cummins EP, Berra E, Comerford KM, Ginouves A, Fitzgerald KT, Seeballuck F, Godson C, Nielsen JE, Moynagh P, Pouyssegur J & Taylor CT (2006) Prolyl hydroxylase-1 negatively regulates I $\kappa$ B kinase- $\beta$ , giving insight into hypoxia-induced NF $\kappa$ B activity. *Proc Natl Acad Sci USA* 103: 18154-18159.
- Dann CE 3rd & Bruick RK (2005) Dioxygenases as O<sub>2</sub>-dependent regulators of the hypoxic response pathway. *Biochem Biophys Res Commun* 338: 639-647.
- de Jong L & Kemp A (1984) Stoichiometry and kinetics of the prolyl 4-hydroxylase partial reaction. *Biochim Biophys Acta* 787: 105-111.
- Dijkstra K, Karvonen P, Pirneskoski A, Koivunen P, Kivirikko KI, Darby NJ, van Straaten M, Scheek RM & Kemmink J (1999) Assignment of <sup>1</sup>H, <sup>13</sup>C and <sup>15</sup>N resonances of the a' domain of protein disulfide isomerase. *J Biomol NMR* 14: 195-196.
- Ellgaard L & Ruddock LW (2005) The human protein disulphide isomerase family: substrate interactions and functional properties. *EMBO Rep* 6: 28-32.
- Engel J & Bächinger HP (2005) Structure, stability and folding of the collagen triple helix. In: *Topics in Current Chemistry*. Springer-Verlag, Berlin, 247: 7-34.
- Epstein AC, Gleadle JM, McNeill LA, Hewitson KS, O'Rourke J, Mole DR, Mukherji M, Metzen E, Wilson MI, Dhanda A, Tian Y-M, Masson N, Hamilton DL, Jaakkola P, Barstead R, Hodgkin J, Maxwell PH, Pugh CW, Schofield CJ & Ratcliffe PJ (2001) *C. elegans* EGL-9 and mammalian homologs define a family of dioxygenases that regulate HIF by prolyl hydroxylation. *Cell* 107: 43-54.
- Eriksson M, Myllyharju J, Tu H, Hellman M & Kivirikko KI (1999) Evidence for 4-hydroxyproline in viral proteins. Characterization of a viral prolyl 4-hydroxylase and its peptide substrates. *J Biol Chem* 274: 22131-22134.
- 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.
- Fedele AO, Whitelaw ML & Peet DJ (2002) Regulation of gene expression by the hypoxia-inducible factors. *Mol Interv* 2: 229-243.
- Fietzek PP, Wendt P, Kell I & Kühn K (1972) The covalent structure of collagen: amino acid sequence of I-CB3 from calf skin collagen. *FEBS Lett* 26: 74-76.

- 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.
- Gill AC, Ritchie MA, Hunt LG, Steane SE, Davies KG, Bocking SP, Rhie AG, Bennett AD & Hope J (2000) Post-translational hydroxylation at the N-terminus of the prion protein reveals presence of PPII structure in vivo. *EMBO J* 19: 5324-5331.
- Gryder RM, Lamon M & Adams E (1975) Sequence position of 3-hydroxyproline in basement membrane collagen. Isolation of glycyL-3-hydroxyprolyl-4-hydroxyproline from swine kidney. *J Biol Chem* 250: 2470-2474.
- Hanauske-Abel HM (1991) Prolyl 4-hydroxylase, a target enzyme for drug development. Design of suppressive agents and the *in vitro* effects of inhibitors and proinhibitors. *J Hepatol* 3: 8-15.
- Hanauske-Abel HM & Günzler V (1982) A stereochemical concept for the catalytic mechanism of prolylhydroxylase: applicability to classification and design of inhibitors. *J Theor Biol* 94: 421-455.
- 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.
- 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  $\alpha$ -subunit isoform: formation of an  $\alpha_2\beta_2$  tetramer with the protein disulfide-isomerase/ $\beta$  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  $\alpha$  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  $\alpha$ -subunit of human prolyl 4-hydroxylase: the complete cDNA-derived amino acid sequence and evidence for alternative splicing of RNA transcripts. *Proc Natl Acad Sci USA* 86: 4392-4396.
- Hewitson KS, McNeill LA, Riordan MV, Tian YM, Bullock AN, Welford RW, Elkins JM, Oldham NJ, Bhattacharya S, Gleadle JM, Ratcliffe PJ, Pugh CW & Schofield CJ (2002) Hypoxia-inducible factor (HIF) asparagine hydroxylase is identical to factor inhibiting HIF (FIH) and is related to the cupin structural family. *J Biol Chem* 277: 26351-26355.
- Hewitson KS, Lienard BM, McDonough MA, Clifton IJ, Butler D, Soares AS, Oldham NJ, McNeill LA & Schofield CJ (2007) Structural and mechanistic studies on the inhibition of the hypoxia-inducible transcription factor hydroxylases by tricarboxylic acid cycle intermediates. *J Biol Chem* 282: 3293-3301.

- 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  $\alpha$ -like peptides. *J Biol Chem* 277: 23965-23971.
- 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.
- Hill KL, Harfe BD, Dobbins CA & L'Hernault SW (2000) *dpy-18* encodes an  $\alpha$ -subunit of prolyl-4-hydroxylase in *Caenorhabditis elegans*. *Genetics* 155: 1139-1148.
- Hirsilä M, Koivunen P, Günzler V, Kivirikko KI & Myllyharju J (2003) Characterization of the human prolyl 4-hydroxylases that modify the hypoxia-inducible factor. *J Biol Chem* 278: 30772-30780.
- Hirsilä M, Koivunen P, Xu L, Seeley T, Kivirikko KI & Myllyharju J (2005) Effect of desferrioxamine and metals on the hydroxylases in the oxygen sensing pathway. *FASEB J* 19: 1308-1310.
- 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  $\alpha$  subunit. *Gene* 149: 391-392.
- Huang J, Zhao Q, Mooney SM & Lee FS (2002) Sequence determinants in hypoxia-inducible factor-1 $\alpha$  for hydroxylation by the prolyl hydroxylases PHD1, PHD2, and PHD3. *J Biol Chem* 277: 39792-39800.
- Huang LE, Pete EA, Schau M, Milligan J & Gu J (2002) Leu-574 of HIF-1 $\alpha$  is essential for the von Hippel-Lindau (VHL)-mediated degradation pathway. *J Biol Chem* 277: 41750-41755.
- Inouye K, Kobayashi Y, Kyogoku Y, Kishida Y, Sakakibara S & Prockop DJ (1982) Synthesis and physical properties of (hydroxyproline-proline-glycine)<sub>10</sub>: hydroxyproline in the X-position decreases the melting temperature of the collagen triple helix. *Arch Biochem Biophys* 219: 198-203.
- Isaacs JS, Jung YJ, Mole DR, Lee S, Torres-Cabala C, Chung YL, Merino M, Trepel J, Zbar B, Toro J, Ratcliffe PJ, Linehan WM & Neckers L (2005) HIF overexpression correlates with biallelic loss of fumarate hydratase in renal cancer: novel role of fumarate in regulation of HIF stability. *Cancer Cell* 8: 143-153.
- 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 $\alpha$  targeted for VHL-mediated destruction by proline hydroxylation: implications for O<sub>2</sub> sensing. *Science* 292: 464-468.
- Jaakkola P, Mole DR, Tian Y-M, Wilson MI, Gielbert J, Gaskell SJ, Kriegsheim A, Hestrestreit HF, Mukherji M, Schofield CJ, Maxwell PH, Pugh CW & Ratcliffe PJ (2001) Targeting of HIF- $\alpha$  to the von Hippel-Lindau ubiquitylation complex by O<sub>2</sub>-regulated prolyl hydroxylation. *Science* 292: 468-472.
- Järnum S, Kjellman C, Darabi A, Nilsson I, Edvardsen K & Åman P (2004) LEPREL1, a novel ER and Golgi resident member of the Leprecan family. *Biochem Biophys Res Commun* 317: 342-351.
- Jenkins CL & Raines RT (2002) Insights on the conformational stability of collagen. *Nat Prod Rep* 19: 49-59.
- Jenkins CL, Bretscher LE, Guzei IA & Raines RT (2003) Effect of 3-hydroxyproline residues on collagen stability. *J Am Chem Soc* 125: 6422-6427.
- Jiang BH, Rue E, Wang GL, Roe R & Semenza GL (1996) Dimerization, DNA binding, and transactivation properties of hypoxia-inducible factor 1. *J Biol Chem* 271: 17771-17778.
- Jimenez S, Harsch M & Rosenbloom J (1973) Hydroxyproline stabilizes the triple helix of chick tendon collagen. *Biochem Biophys Res Commun* 52: 106-114.
- John DCA & 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.
- Kaelin WG Jr (2005a) Proline hydroxylation and gene expression. *Annu Rev Biochem* 74: 115-128.
- Kaelin WG Jr (2005b) The von Hippel-Lindau, HIF hydroxylation, and oxygen sensing. *Biochem Biophys Res Commun* 338: 627-638.
- Kallio PJ, Okamoto K, O'Brien S, Carrero P, Makino Y, Tanaka H & Poellinger L (1998) Signal transduction in hypoxic cells: inducible nuclear translocation and recruitment of the CBP/p300 coactivator by the hypoxia-inducible factor-1 $\alpha$ . *EMBO J* 17: 6573-6586.
- 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.
- Kaul SC, Sugihara T, Yoshida A, Nomura H & Wadhwa R (2000) Gros1, a potential growth suppressor on chromosome 1: its identity to basement membrane-associated proteoglycan, leprecan. *Oncogene* 19: 3576-3583.
- Ke Q & Costa M (2006) Hypoxia-inducible factor-1 (HIF-1). *Mol Pharmacol* 70: 1469-1480.

- Kefalides NA (1973) Structure and biosynthesis of basement membranes. *Int Rev Connect Tissue Res* 6: 63-104.
- 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.
- 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.
- Khan Z, Michalopoulos GK & Stolz DB (2006) Peroxisomal localization of hypoxia-inducible factors and hypoxia-inducible factor regulatory hydroxylases in primary rat hepatocytes exposed to hypoxia-reoxygenation. *Am J Pathol* 169: 1251-1269.
- 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.
- Kielty CM & Grant ME (2002) The collagen family: structure, assembly, and organization in the extracellular matrix. In: Royce PM & Steinmann B (eds) *Connective Tissue and Its Heritable Disorders*. Wiley-Liss Inc, New York, 159-221.
- Kivirikko KI & Myllyharju J (1998) Prolyl 4-hydroxylases and their protein disulfide isomerase subunit. *Matrix Biol* 16: 357-368.
- Kivirikko KI & Myllylä R (1982) Posttranslational enzymes in the biosynthesis of collagen: intracellular enzymes. *Methods Enzymol* 82: 245-304.
- Kivirikko KI, Myllylä R & Pihlajaniemi T (1992) Hydroxylation of proline and lysine residues in collagens and other animal and plant proteins. In: Harding JJ & Crabbe 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.
- Köditz J, Nesper J, Wottawa M, Stiehl DP, Camenisch G, Franke C, Myllyharju J, Wenger RH, Katschinski DM (2007) Oxygen-dependent ATF-4 stability is mediated by the PHD3 oxygen sensor. *Blood*, in press.
- Koide T & Nagata K (2005) Collagen biosynthesis. In: Brickmann J, Notbohm H, Müller PK (eds) *Topics in Current Chemistry*. Springer-Verlag, Berlin, 247: 85-114.

- Koivu J, Myllylä R, Helaakoski T, Pihlajaniemi T, Tasanen K & Kivirikko KI (1987) A single polypeptide acts both as the  $\beta$  subunit of prolyl 4-hydroxylase and as a protein disulfide-isomerase. *J Biol Chem* 262: 6447-6449.
- Koivunen P, Hirsilä M, Kivirikko KI & Myllyharju J (2006) The length of peptide substrates has a marked effect on hydroxylation by the hypoxia-inducible factor prolyl 4-hydroxylases. *J Biol Chem* 281: 28712-28720.
- Koivunen P, Hirsilä M, Remes AM, Hassinen IE, Kivirikko KI & Myllyharju J (2007) Inhibition of hypoxia-inducible factor (HIF) hydroxylases by citric acid cycle intermediates: possible links between cell metabolism and stabilization of HIF. *J Biol Chem* 282: 4524-4532.
- 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.
- Kresina TF & Miller EJ (1979) Isolation and characterization of basement membrane collagen from human placental tissue. Evidence for the presence of two genetically distinct collagen chains. *Biochemistry* 18: 3089-3097.
- 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.
- Kuznetsova AV, Meller J, Schnell PO, Nash JA, Ignacak ML, Sanchez Y, Conaway JW, Conaway RC & Czyzyk-Krzeska MF (2003) von Hippel-Lindau protein binds hyperphosphorylated large subunit of RNA polymerase II through a proline hydroxylation motif and targets it for ubiquitination. *Proc Natl Acad Sci USA* 100: 2706-2711.
- Labuda M, Morissette J, Ward LM, Rauch F, Lalic L, Roughley PJ & Glorieux FH (2002) Osteogenesis imperfecta type VII maps to the short arm of chromosome 3. *Bone* 31: 19-25.
- Lahiri S, Roy A, Baby SM, Hoshi T, Semenza GL & Prabhakar NR (2006) Oxygen sensing in the body. *Prog Biophys Mol Biol* 91: 249-286.
- Lamberg A, Pihlajaniemi T & Kivirikko KI (1995) Site-directed mutagenesis of the  $\alpha$  subunit of human prolyl 4-hydroxylase. Identification of three histidine residues critical for catalytic activity. *J Biol Chem* 270: 9926-9931.
- Lando D, Peet D, Gorman JJ, Whelan DA, Whitelaw ML & Bruick RK (2002a) FIH-1 is an asparaginyl hydroxylase enzyme that regulates the transcriptional activity of hypoxia inducible factor. *Genes Dev* 16: 1466-1471.
- Lando D, Peet D, Whelan DA, Gorman JJ & Whitelaw ML (2002b) Asparagine hydroxylation of the HIF transactivation domain: a hypoxic switch. *Science* 295: 858-862.
- Lee S, Nakamura E, Yang H, Wei W, Linggi MS, Sajan MP, Farese RV, Freeman RS, Carter BD, Kaelin WG Jr & Schlisio S (2005) Neuronal apoptosis linked to EglN3 prolyl hydroxylase and familial pheochromocytoma genes: developmental culling and cancer. *Cancer Cell* 8: 155-167

- Li D, Hirsilä M, Koivunen P, Brenner MC, Xu L, Yang C, Kivirikko KI & Myllyharju J (2004) Many amino acid substitutions in a hypoxia-inducible transcription factor (HIF)-1 $\alpha$ -like peptide cause only minor changes in its hydroxylation by the HIF prolyl 4-hydroxylases: substitution of 3,4-dehydroproline or azetidine-2-carboxylic acid for the proline leads to a high rate of uncoupled 2-oxoglutarate decarboxylation. *J Biol Chem* 279: 55051-55059.
- Lieb ME, Menzies K, Moschella MC, Ni R & Taubman MB (2002) Mammalian EGLN genes have distinct patterns of mRNA expression and regulation. *Biochem Cell Biol* 80: 421-426.
- Liu Q, Berchner-Pfannschmidt U, Möller U, Brecht M, Wotzlaw C, Acker H, Jungermann K & Kietzmann T (2004) A Fenton reaction at the endoplasmic reticulum is involved in the redox control of hypoxia-inducible gene expression. *Proc Natl Acad Sci USA* 101: 4302-4307.
- Mahon PC, Hirota K & Semenza GL (2001) FIH-1: a novel protein that interacts with HIF-1 $\alpha$  and VHL to mediate repression of HIF-1 transcriptional activity. *Genes Dev* 15: 2675-2686.
- Majamaa K, Günzler V, Hanauske-Abel HM, Myllylä R & Kivirikko KI (1986) Partial identity of the 2-oxoglutarate and ascorbate binding sites of prolyl 4-hydroxylase. *J Biol Chem* 261: 7819-7823.
- Majamaa K, Hanauske-Abel HM, Günzler V & Kivirikko KI (1984) The 2-oxoglutarate binding site of prolyl 4-hydroxylase. Identification of distinct subsites and evidence for 2-oxoglutarate decarboxylation in a ligand reaction at the enzyme-bound ferrous ion. *Eur J Biochem* 138: 239-245.
- Masson N & Ratcliffe PJ (2003) HIF prolyl and asparaginyl hydroxylases in the biological response to intracellular O<sub>2</sub> levels. *J Cell Sci* 116: 3041-3049.
- Masson N, Willam C, Maxwell PH, Pugh CW & Ratcliffe PJ (2001) Independent function of two destruction domains in hypoxia-inducible factor- $\alpha$  chains activated by prolyl hydroxylation. *EMBO J* 20: 5197-5206.
- McDonough MA, Li V, Flasman 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.
- 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.
- Metzen E & Ratcliffe PJ (2004) HIF hydroxylation and cellular oxygen sensing. *Biol Chem* 385: 223-230.
- Metzen E, Berchner-Pfannschmidt U, Stengel P, Marxsen JH, Stolze I, Klinger M, Huang WQ, Wotzlaw C, Hellwig-Burgel T, Jelkmann W, Acker H & Fandrey J (2003a) Intracellular localisation of human HIF-1 $\alpha$  hydroxylases: implications for oxygen sensing. *J Cell Sci* 116: 1319-1326.

- Metzen E, Zhou J, Jelkmann W, Fandrey J & Brüne B (2003b) Nitric oxide impairs normoxic degradation of HIF-1 $\alpha$  by inhibition of prolyl hydroxylases. *Mol Biol Cell* 14: 3470-3481.
- Miller EJ, Harris ED Jr, Chung E, Finch JE Jr, McCroskery PA & Butler WT (1976) Cleavage of type II and III collagens with mammalian collagenase: site of cleavage and primary structure at the NH<sub>2</sub>-terminal portion of the smaller fragment released from both collagens. *Biochemistry* 15: 787-792.
- Mizuno K, Hayashi T, Peyton DH, Bächinger HP (2004) The peptides acetyl-(Gly-3(S)Hyp-4(R)Hyp)<sub>10</sub>-NH<sub>2</sub> and acetyl-(Gly-Pro-3(S)Hyp)<sub>10</sub>-NH<sub>2</sub> do not form a collagen triple helix. *J Biol Chem* 279: 282-287.
- Morello R, Bertin TK, Chen Y, Hicks J, Tonachini L, Monticone M, Castagnola P, Rauch F, Glorieux FH, Vranka J, Bächinger HP, Pace JM, Schwarze U, Byers PH, Weis M, Fernandes RJ, Eyre DR, Yao Z, Boyce BF & Lee B (2006) CRTAP is required for prolyl 3- hydroxylation and mutations cause recessive osteogenesis imperfecta. *Cell* 127: 291-304.
- 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, Berlin, 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 & Kivirikko KI (2001) Collagens and collagen-related diseases. *Ann Med* 33: 7-21.
- 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.
- Myllylä R, Tuderman L & Kivirikko KI (1977) Mechanism of the prolyl hydroxylase reaction. 2. Kinetic analysis of the reaction sequence. *Eur J Biochem* 80: 349-357.
- Nietfeld JJ, Van der Kraan I & Kemp A (1981) Dissociation and reassociation of prolyl 4-hydroxylase subunits after cross-linking of monomers. *Biochem Biophys Acta* 661: 21-27.
- Nissi R, Autio-Harmainen H, Marttila P, Sormunen R & Kivirikko KI (2001) Prolyl 4-hydroxylase isoenzymes I and II have different expression patterns in several human tissues. *J Histochem Cytochem* 49: 1143-1153.

- Nokelainen M, Nissi R, Kukkola L, Helaakoski T & Myllyharju J (2001) Characterization of the human and mouse genes for the  $\alpha$  subunit of type II prolyl 4-hydroxylase. Identification of a previously unknown alternatively spliced exon and its expression in various human tissues. *Eur J Biochem* 268: 5300-5309.
- Oehme F, Ellinghaus P, Kolkhof P, Smith TJ, Ramakrishnan S, Hutter J, Schramm M & Flamme I (2002) Overexpression of PH-4, a novel putative proline 4-hydroxylase, modulates activity of hypoxia-inducible transcription factors. *Biochem Biophys Res Commun* 296: 343-349.
- 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  $\alpha$ -subunit of prolyl 4-hydroxylase to human chromosome region 10q21.3-23.1. *Am J Hum Genet* 45: 829-834.
- Passoja K, Myllyharju J, Pirskanen A & Kivirikko KI (1998) Identification of arginine-700 as the residue that binds the C-5 carboxyl group of 2-oxoglutarate in human lysyl hydroxylase 1. *FEBS Lett* 434: 145-148.
- 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.
- Pihlajaniemi T, Helaakoski T, Tasanen K, Myllylä R, Huhtala ML, Koivu J & Kivirikko KI (1987) Molecular cloning of the  $\beta$ -subunit of human prolyl 4-hydroxylase. This subunit and protein disulphide isomerase are products of the same gene. *EMBO J* 6: 643-649.
- 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.
- Rauch F & Glorieux FH (2004) Osteogenesis imperfecta. *Lancet* 363: 1377-1385.
- Ricard-Blum S & Ruggiero F (2005) The collagen superfamily: from the extracellular matrix to the cell membrane. *Pathol Biol (Paris)* 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.
- Risteli J, Tryggvason K & Kivirikko KI (1977) Prolyl 3-hydroxylase: partial characterization of the enzyme from rat kidney cortex. *Eur J Biochem* 73: 485-492.

- Rodgers UR & Weiss AS (2005) Cellular interactions with elastin. *Pathol Biol (Paris)* 53: 390-398.
- Schofield CJ & Ratcliffe PJ (2005) Signalling hypoxia by HIF hydroxylases. *Biochem Biophys Res Commun* 338: 617-626.
- Schofield CJ & Zhang Z (1999) Structural and mechanistic studies on 2-oxoglutarate-dependent oxygenases and related enzymes. *Curr Opin Struct Biol* 9: 722-731.
- Schumacher MA, Mizuno K & Bächinger HP (2006) The crystal structure of a collagen-like polypeptide with 3(S)-hydroxyproline residues in the Xaa position forms a standard 7/2 collagen triple helix. *J Biol Chem* 281: 27566-27574.
- Selak MA, Armour SM, MacKenzie ED, Boulahbel H, Watson DG, Mansfield KD, Pan Y, Simon MC, Thompson CB & Gottlieb E (2005) Succinate links TCA cycle dysfunction to oncogenesis by inhibiting HIF- $\alpha$  prolyl hydroxylase. *Cancer Cell* 7: 77-85.
- Semenza GL (2006) Regulation of physiological responses to continuous and intermittent hypoxia by hypoxia-inducible factor 1. *Exp Physiol* 91: 803-806.
- 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.
- Siddiq A, Aminova LA & Ratan RR (2007) Hypoxia inducible factor prolyl 4-hydroxylase enzymes: center stage in the battle against hypoxia, metabolic compromise and oxidative stress. *Neurochem Res* 32: 931-946.
- Smith MC, Burns N, Sayers JR, Sorrell JA, Casjens SR & Hendrix RW (1998) Bacteriophage collagen. *Science* 279: 1834.
- Srinivas V, Zhang LP, Zhu XH & Caro J (1999) Characterization of an oxygen/redox-dependent degradation domain of hypoxia-inducible factor  $\alpha$  (HIF- $\alpha$ ) proteins. *Biochem Biophys Res Commun* 260: 557-561.
- Sumper M & Hallmann A (1998) Biochemistry of the extracellular matrix of *Volvox*. *Int Rev Cytol* 180: 51-85.
- Takeda K, Ho VC, Takeda H, Duan LJ, Nagy A & Fong GH (2006) Placental but not heart defects are associated with elevated hypoxia-inducible factor  $\alpha$  levels in mice lacking prolyl hydroxylase domain protein 2. *Mol Cell Biol* 26: 8336-8346.
- Tanaka M, Shibata H & Uchida T (1980) A new prolyl hydroxylase acting on poly-L-proline, from suspension cultured cells of *Vinca rosea*. *Biochem Biophys Acta* 61: 188-198.
- Taylor MS (2001) Characterization and comparative analysis of the EGLN gene family. *Gene* 275: 125-132.
- 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.
- Tonachini L, Morello R, Monticone M, Skaug J, Scherer SW, Cancedda R & Castagnola P (1999) cDNA cloning, characterization and chromosome mapping of the gene encoding human cartilage associated protein (CRTAP). *Cytogenet Cell Genet* 87: 191-194.

- Tryggvason K, Majamaa K, Risteli J & Kivirikko KI (1979) Partial purification and characterization of chick-embryo prolyl 3-hydroxylase. *Biochem J* 183: 303-307.
- Tryggvason K, Risteli J & Kivirikko KI (1977) Separation of prolyl 3-hydroxylase and 4-hydroxylase activities and the 4-hydroxyproline requirement for synthesis of 3-hydroxyproline. *Biochem Biophys Res Commun* 76: 275-281.
- Tuderman L, Myllylä R & Kivirikko KI (1977) Mechanism of the prolyl hydroxylase reaction. I. Role of co-substrates. *Eur J Biochem* 80: 341-348.
- Van Damme EJ, Barre A, Rougé P & Peumans WJ (2004) Cytoplasmic/nuclear plant lectins: a new story. *Trends Plant Sci* 9: 484-489.
- Van Den Diepstraten C, Papay K, Bolender Z, Brown A & Pickering JG (2003) Cloning of a novel prolyl 4-hydroxylase subunit expressed in the fibrous cap of human atherosclerotic plaque. *Circulation* 108: 508-511.
- Veijola J, Annunen P, Koivunen P, Page AP, Pihlajaniemi T & Kivirikko KI (1996) Baculovirus expression of two protein disulphide isomerase isoforms from *Caenorhabditis elegans* and characterization of prolyl 4-hydroxylases containing one of these polypeptides as their  $\beta$  subunit. *Biochem J* 317: 721-729.
- Veijola J, Koivunen P, Annunen P, Pihlajaniemi T & Kivirikko KI (1994) Cloning, baculovirus expression, and characterization of the  $\alpha$  subunit of prolyl 4-hydroxylase from the nematode *Caenorhabditis elegans*. This  $\alpha$  subunit forms an active  $\alpha\beta$  dimer with the human protein disulfide isomerase/ $\beta$  subunit. *J Biol Chem* 269: 26746-26753.
- Veit G, Kobbe B, Keene DR, Paulsson M, Koch M & Wagener R (2006) Collagen XXVIII, a novel VWA-domain-containing protein with many imperfections in the collagenous domain. *J Biol Chem* 281: 3494-3504.
- Villar D, Vara-Vega A, Landázuri MO & del Peso L (2007) Identification of a region on HIF prolyl 4-hydroxylases that determines their specificity for the oxygen degradation domains. *Biochem J*, in press.
- Vranka JA, Sakai LY & Bächinger HP (2004) Prolyl 3-hydroxylase 1, enzyme characterization and identification of a novel family of enzymes. *J Biol Chem* 279: 23615-23621.
- 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.
- Wang GL, Jiang BH, Rue EA & Semenza GL (1995) Hypoxia-inducible factor 1 is a basic-helix-loop-helix-PAS heterodimer regulated by cellular O<sub>2</sub> tension. *Proc Natl Acad Sci USA* 92: 5510-5514.
- Wang J & Pantopoulos K (2005) The pathway for IRP2 degradation involving 2-oxoglutarate-dependent oxygenase(s) does not require the E3 ubiquitin ligase activity of pVHL. *Biochim Biophys Acta* 1743: 79-85.

- Wassenhove-McCarthy DJ & McCarthy KJ (1999) Molecular characterization of a novel basement membrane-associated proteoglycan, leprecan. *J Biol Chem* 274: 25004-25017.
- Wijffels GL, Panaccio M, Salvatore L, Wilson L, Walker ID & Spithill TW (1994) The secreted cathepsin L-like proteinases of the trematode, *Fasciola hepatica*, contain 3-hydroxyproline residues. *Biochem J* 299: 781-790.
- Wilkinson B & Gilbert HF (2004) Protein disulfide isomerase. *Biochim Biophys Acta* 1699: 35-44.
- 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.
- Winter AD, Keskiäho K, Kukkola L, McCormack G, Felix MA, 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* 26: 382-395.
- 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.
- Wojtaszek P, Smith CG & Bolwell GP (1999) Ultrastructural localisation and further biochemical characterisation of prolyl 4-hydroxylase from *Phaseolus vulgaris*: comparative analysis. *Int J Biochem Cell Biol* 31: 463-477.
- Xu Y, Keene DR, Bujnicki JM, Höök M & Lukomski S (2002) Streptococcal Sc11 and Sc12 proteins form collagen-like triple helices. *J Biol Chem* 277: 27312-27318.
- Yu F, White SB, Zhao Q & Lee FS (2001) HIF-1 $\alpha$  binding to VHL is regulated by stimulus-sensitive proline hydroxylation. *Proc Natl Acad Sci USA* 98: 9630-9635.
- 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.

## Original articles

- I Koivunen P, Tiainen P, Hyvärinen J, Williams KE, Sormunen R, Klaus SJ, Kivirikko KI & Myllyharju J (2007) An endoplasmic reticulum transmembrane prolyl 4-hydroxylase is induced by hypoxia and acts on hypoxia-inducible factor  $\alpha$ . *J Biol Chem* 282: 30544-30552.
- II 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.
- III Tiainen P, Pasanen A, Sormunen R & Myllyharju J (2007) Characterization of recombinant human prolyl 3-hydroxylase isoenzyme 2, a basement membrane enzyme. Submitted.

The permission of the copyright owner American Society for Biochemistry and Molecular Biology to reproduce the original articles I and II has been granted.

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



943. Arpiainen, Satu (2007) Transcriptional regulation of the hepatic cytochrome *P450 2a5* gene
944. Päiväläinen-Jalonen, Satu (2007) Expression and stability of myelin-associated elements
945. Tiainen, Johanna (2007) Bioresorbable plain and ciprofloxacin-releasing self-reinforced PLGA 80/20 implants' suitability for craniofacial surgery. Histological and mechanical assessment
946. Aaltonen, Vesa (2007) PKC and neurofibromin in the molecular pathology of urinary bladder carcinoma. The effect of PKC inhibitors on carcinoma cell junctions, movement and death
947. Kuvaja, Paula (2007) The prognostic role of matrix metalloproteinases MMP-2 and -9 and their tissue inhibitors TIMP-1 and -2 in primary breast carcinoma
948. Siira, Virva (2007) Vulnerability signs of mental disorders in adoptees with genetic liability to schizophrenia and their controls measured with Minnesota Multiphasic Personality Inventory
949. Komulainen, Silja (2007) Effect of antihypertensive drugs on blood pressure during exposure to cold. Experimental study in normotensive and hypertensive subjects
950. Anttonen, Vuokko (2007) Laser fluorescence in detecting and monitoring the progression of occlusal dental caries lesions and for screening persons with unfavourable dietary habits
951. Torppa, Kaarina (2007) Managerialismi suomalaisen julkisen erikoissairaanhoidon johtamisessa. Tutkimus yksityissektorin johtamisoppien soveltamisesta neljässä yliopistollisessa sairaanhoitopiirissä ja arvio managerialismin soveltuvuudesta julkisen erikoissairaanhoidon uudistamiseen
952. Hokkanen, Eero (2007) Neurologian kehitysvaiheita Oulussa ja Pohjois-Suomessa
953. Lahtinen, Jarmo (2007) Predictors of immediate outcome after coronary artery bypass surgery
954. Železnik, Danica (2007) Self-care of the home-dwelling elderly people living in Slovenia
955. Löfgren, Eeva (2007) Effects of epilepsy and antiepileptic medication on reproductive function
956. Jääliinoja, Juha (2007) The structure and function of normal and mutated collagen IX

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 978-951-42-8672-8 (Paperback)

ISBN 978-951-42-8673-5 (PDF)

ISSN 0355-3221 (Print)

ISSN 1796-2234 (Online)

