DEVELOPMENT OF TRANSGENIC BARLEY EXPRESSING HUMAN TYPE I COLLAGEN

By

CLAUDIA E. OSORIO

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

MASTER OF SCIENCE IN CROP SCIENCE

WASHINGTON STATE UNIVERSITY
Department of Crop and Soil Sciences

DECEMBER 2004

To the Faculty of Washington State University:

The members of the Committee appointed to examine the dissertation of CLAUDIA E. OSORIO find it satisfactory and recommend that it be accepted.

_____

Chair

_____

_____

Acknowledgements

I would like to thank professor Diter von Wettstein for providing me with the opportunity of work with him and his group, for his patience, advice and experience given through the completion of this thesis. I really appreciate the opportunity to work with a kind person and a distinguished scientist. I wish to thank all the people working in our group for the constant support during these years, specially I would like to express my sincere thanks to Dr. Gamini Kannangara for his help, advise and optimism during the days spent in the lab.

My warmest thanks belong to my family and friends for their encouragement and unconditional belief in me. Above all, thanks to Jaime, Paulina, Felipe and Pablo for give me every day a reason to be grateful, I owe them my greatest gratitude because of their love, confidence and support, which has been essential in bringing my thesis to a good end.

DEVELOPMENT OF TRANSGENIC BARLEY EXPRESSING HUMAN TYPE I COLLAGEN

Abstract

by Claudia E. Osorio, M.S.
Washington State University
December 2004

Chair: Diter von Wettstein

Collagen I is the main protein found in the extracellular matrix. It consists of a heterotrimeric triple helix that aggregates to form a fibrillar structure. Eight specific enzymes are needed for correct folding and secretion of the protein to the extracellular space. The need to find a non-animal source for the production of collagen has led to the development of expression systems that are able to produce recombinant collagen. The main objective of this dissertation was to produce transgenic barley grains expressing the genes needed for production of hydroxylated procollagen in the endosperm. To prepare for reaching this objective, it was necessary to codon optimize the gene and express it in an eucaryotic system like *Pichia pastoris* for verification that the correct protein product was synthesized. Then *Agrobacterium tumefaciens* mediated transformation of immature zygotic embryos of barley could be initiated with the appropriate vectors. Accordingly, vectors were constructed for expression in

Pichia pastoris and barley. Homotrimeric procollagen I and prolyl 4-hydroxylase were successfully expressed in *Pichia pastoris*, as shown by SDS-PAGE, Western blot and ELISA. Barley was transformed by *Agrobacterium tumefaciens* with vectors carrying the genes for collagen and both subunits of prolyl 4-hydroxylase. Green plants were selected with the aid of bialaphos resistance and PCR-tested with specific primers for the internal part of (α)1 collagen I chain.

Table of Contents

List of Tables

List of Figures

Abbreviations

Table 1: Abbreviations used

Abbreviation	
AOX	Alcohol oxidase
CIM	Callus induction medium
DMSO	Dimethyl sulfoxide
EDTA	Disodium ethylene diamine tetraacetate
ELISA	Enzyme-linked immunosorbent assay
FACIT	Fibril associated collagens with interrupted triple helices
IgG	Immunoglobulin G
LB	Luria Bertani
LH	Lysyl hydroxylase
L-PPT	L-phosphinothricin
P3H	Prolyl 3-hydroxylase
P4H	Prolyl 4-hydroxylase
PAT	Phosphinothricin acetyl transferase
PCR	Polymerase chain reaction
PDI	Protein disulfide isomerase
PEG	Polyethylene glycol
PHO1	*Schizosaccharamyces pombe* acid phosphatase 1
RGM	Root generation medium
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SGM	Shoot generation medium
SOE	Splicing by overlap extension
TBS	Tris- buffered saline
YPD	Yeast peptone dextrose

To Jaime, Paulina, Felipe and Pablo

Chapter 1

Introduction

1.1 General introduction

Production of proteins with pharmaceutical purposes is a developing industry, but the extraction of proteins from their natural source involves risks to health and, depending on the protein, these processes are sometimes inefficient and expensive. Moreover, some proteins are not available naturally. To counteract these aspects, different recombinant systems have been developed during recent years, which include expression of proteins in *E. coli*, yeast, insect and mammalian cell cultures, as well as transgenic animals and plants [*Hood,* 2002*; Ma et al.,* 2003*; Fischer et al.,* 2003*; Twyman et al.,* 2003].

Each of these systems has disadvantages [*Hood,* 2002]: bacteria, fungi and yeast require a high initial input of money, moreover; in the case of bacteria [*Twyman et al.,* 2003], many proteins fail to fold correctly and are either degraded (resulting in low or no yield) or accumulate as insoluble inclusion bodies. In the case of fungi, hyper-

glycosylation of proteins [*Hood,* 2002*; Streatfield et al.,* 2003] can lead to undesired products. Animal cell culture and transgenic animal hosts have the advantage that the expression of the protein is similar to the original provenance, but also involve a high risk of viral diseases, are expensive and only cost effective for highly valuable products [*Hood,* 2002*; Olsen et al.,* 2003].

In the last few years, plants were added to systems for production of recombinant proteins, with some successful approaches that make this alternative a viable approach. At this point, several plant-derived biopharmaceutical proteins are reaching the stage of clinical trials, including antibodies, vaccines, human blood products, hormones and growth regulators [*Fischer et al.,* 2003*; Peterson and Arntzen,* 2004*; Twyman et al.,* 2003]. The advantages of plant products include lower costs compared with microbial or animal cells, stability of protein products in storage tissues such as seeds and the possibility of direct addition of plant material to industrial processes [*Hood,* 2002]. Moreover, plants seems to have the ability for correctly folding and assembling foreign proteins, as the result of a high conservation of protein synthesis pathways [*Twyman et al.,* 2003].

Recombinant proteins expressed in plants

The first mammalian protein successfully expressed and correctly folded in a transgenic plant was reported by Sijmons and collaborators in 1990 [*Sijmons et al.,* 1990]. These authors achieved secretion of correctly processed human serum albumin from tobacco

cells. Since then, many other proteins have been successfully expressed in different crops.

In the late 80's, assembly of full length serum IgG was realized in tobacco by crossing transgenic plants that expressed, respectively, gamma or kappa chains. The functional antibodies accumulated up to 1.3% of total leaf protein content [*Hiatt et al.,* 1989]. Following the same procedure of crosses among transgenic lines expressing genes for the heavy and light chains, it was possible to synthesize and correctly assemble secretory immunoglobulin A, a process that involved crosses between four transgenic lines of tobacco plants, each of them expressing a different gene necessary for assembly of the dimeric form [*Ma et al.,* 1995]. Also, the immune response in humans against hepatitis B virus was demonstrated after ingestion of transgenic lettuce expressing hepatitis B virus surface antigen [*Kapusta et al.,* 1999].

Early studies were conducted with model species like tobacco, thereby increasing knowledge of gene regulation and protein synthesis in plants. Recently, higher recombinant protein expression levels were obtained using different crop species like corn, canola, rice, soybean and barley, representing a new industry that can satisfy the exploding demand for recombinant proteins [*Fischer et al.,* 2003*; Hood,* 2002*; Twyman et al.,* 2003].

1.2 Collagen

Collagens are extensively used as industrial products with such diverse applications as gelatins in the food industry or in medical applications, tissue engineering and wound sealants [*Friess*, 1998]. This extensive industry uses animal waste as the primary source. Tissues like bone or skin are chemically treated, commonly with alkali, to extract collagen [*Olsen et al.*, 2003]. However, there are concerns about the safety of this material. Some of them involve biocompatibility, viral disease transmission, and the homogeneity of the product, this latter based on the fact that, in nature, collagen I is usually associated with collagen III, and the separation of this two kinds of collagens involves a difficult enzymatic process [*Olsen et al.*, 2003]. Animal-derived collagen can also trigger allergy immune-responses in humans at a rate between 2-4% of the total population [*Lynn et al.*, 2004]. These questions have initiated interest in the development of non-animal resources for production of collagen [*Olsen et al.*, 2003].

1.2.1 Characteristics of collagen I

Collagens form a superfamily of proteins that are the most abundant extracellular matrix molecules. They are expressed in all tissues of the human body and are involved in many important functions that support the architecture, strength and development of tissues and affect cell attachment, proliferation, migration and differentiation. Collagens play an important role in tissue reparation, acting as a network that helps in

the sealing of wounds. On the other hand, an excess can lead to fibrotic diseases in different organs and tissues. The critical role of individual collagens is recognized by the wide spectrum of diseases that have been found as a result of mutations in the molecule coding sequence [*Myllyharju and Kivirikko,* 2001*; van der Rest and Garrone,* 1991*; Kivirikko,* 1993*; Prockop and Kivirikko,* 1995*; Bornstein,* 1980]

The collagen family of proteins includes 27 different types of collagen that aggregate into at least 38 distinct polypeptides and more than 15 additional proteins with collagen-like domains that account for the most important extracellular component in mammalian systems [*Myllyharju and Kivirikko,* 2001*; Keizer-Gunnink et al.,* 2000*; Prockop et al.,* 1998*; Kivirikko,* 1993*; Olsen et al.,* 2003] All collagen molecules form supramolecular aggregates that are stabilized in part by triple helical domain interactions [*Prockop et al.,* 1998*; Fessler and Fessler,* 1978]. Depending on the structure that they form, collagen can be separated into different well-characterized groups (see Table 1.1).

Collagen molecules consist of three polypeptide chains, called α chains, that differ for each collagen type with respect to length and specific amino acid sequence [*Bornstein,* 1980*; Myllyharju and Kivirikko,* 2001*; Bateman et al.,* 1996]. α chains form a left handed helix, in which every third residue comes into the center of the triple helix. For these three chains to go into the center, the third residue must be glycine, since it is the smallest amino acid [*van der Rest and Garrone,* 1991*; Prockop et al.,* 1998*; Kadler et al.,* 1996*; Myllyharju and Kivirikko,* 2001*; Brodsky and Ramshaw,*

1997].

The structural similarities between fibrillar collagen are reflected at the gene level, with highly conserved intron- exon structures. Every chain has the repeated sequence Gly-X-Y in which, about 30% of the time X is proline, and Y is 4- hydroxyproline. These two amino acids provide stability for the triple helix [*Myllyharju and Kivirikko,* 2001*; Brodsky and Ramshaw,* 1997*; Pakkanen et al.,* 2003]. The collagen fibrils in tissues are often heterogeneous, containing more than one collagen type. Type I collagen fibrils can frequently be a mixture with trace amounts of collagen types III, V and XII, providing diversity among collagen types. Moreover, heterogeneity is also provided by alternate splicing of collagen polypeptide chains and the use of alternate promoters [*Kivirikko,* 1993].

Type I collagen is part of the group that forms fibrils. It is found in most connective tissues, especially in dermis, bones, tendon and ligament. It is synthesized in response to injury and in nodules formed as consequence of fibrotic diseases [*van der Rest and Garrone,* 1991*; Keizer-Gunnink et al.,* 2000*; Myllyharju and Kivirikko,* 2001].

Collagen I (homo- and heterotrimer) is formed by two molecules α1(I) and α2(I). The genes that encode for these proteins are named COLA1A1 and COL1A2 and were assigned to chromosome 17q21.3-22 and 7q21.3-q22, respectively [*Bornstein,* 1980*; Myllyharju and Kivirikko,* 2001*; Prockop and Kivirikko,* 1995*; Westerhausen et al.,* 1991].Each chain is made up of 330 Gly- X- Y repeats. At the carboxyl end, a disulfide-bonded globule serves as recognition site for trimer assembly and avoids premature

fibril formation. This globule is separated from the main helix by a C-protease cleavage site that comprises about 30 residues. At the amino terminal end, a second non-triple helical domain constitutes the N-protease cleavage site [*van der Rest and Garrone,* 1991] (see Fig. 1.1).

1.2.2 Biosynthesis of fibril-forming collagen

The biosynthesis of fibril-forming collagens is characterized by the presence of an unusual number of co-translational and post-translational modifications of the polypeptide chains, a unique feature to collagen and collagen-like amino acid sequences [*Bornstein,* 1980*; Kivirikko,* 1998*; Myllyharju and Kivirikko,* 2001].

Processing of these modifications takes place inside the cell, and also after secretion of procollagen into the extracellular space. The synthesis of α chains and their intracellular modifications give rise to the formation of triple helical procollagen molecules. After secretion, extracellular processing will convert these molecules into cross-linked fibers or supramolecular planar aggregates [*Prockop and Kivirikko,* 1995*; Myllyharju and Kivirikko,* 2001*; Bornstein,* 1980*; Kivirikko,* 1980].

Intracellular modifications:

Removal of pre-protein sequences: the pre-pro-α chains are synthesized on membrane-bound polysomes and, while being assembled, pass through the membrane into the cisternae of the rough endoplasmic reticulum. Procollagens have hydrophobic leader se-

quences at their N-terminal amino acid ends that need to be removed at an early stage of processing, thereby transforming the pre-pro-α chain into pro-α chains [*Kivirikko,* 1980*; Bornstein,* 1980*; Myllyharju and Kivirikko,* 2001].

Hydroxylation of prolyl and lysyl residues: these reactions are catalyzed by three different enzymes, namely prolyl 4-hydroxylase, prolyl 3-hydroxylase and lysyl hydroxylase. The three proteins hydroxylate prolyl or lysyl residues and require ferrous ions, 2-oxoglutarate, oxygen and ascorbate [*Kivirikko,* 1980*; Kivirikko and Pihlajaniemi,* 1998*; Vranka et al.,* 2004](for reactions, see Fig. 1.2).

Prolyl 4-hydroxylase is located in the lumen of the endoplasmic reticulum [*Myllyharju,* 2003*; Bassuk and Berg,* 1989]. In vertebrates, it has the form of a tetramer $\alpha 2\beta 2$, with a molecular weight of 240kDa and consists of two inactive monomers, namely α and β subunits, the later being identical to protein disulfide isomerase (PDI) [*Pihlajaniemi et al.,* 1987].

The human-α subunit consists of 517 amino acids and a signal peptide of 17 residues, giving rise to a protein of 63kDa that contains the catalytic site for hydroxylation [*Kivirikko,* 1980*; Kivirikko and Pihlajaniemi,* 1998]. PDI (55kDa) confers solubility to the α subunit and also catalyzes the thiol/disulfide exchanges in proteins that results in the disulfide bonds essential for protein stability. As a third function, PDI acts as a chaperone that binds the procollagen and inhibits the aggregation of the procollagen chains during translation [*Wilson et al.,* 1998]. When PDI is acting as the β subunit of the prolyl 4-hydroxylase tetramer, it retains up to 50% of its protein

disulfide isomerase activity [*Pihlajaniemi et al.,* 1987*; Vuorela et al.,* 1997*; Kivirikko and Pihlajaniemi,* 1998*; Myllyharju and Kivirikko,* 2001*; Myllyharju,* 2003].

The minimum requirement for interaction with prolyl 4- hydroxylase is fulfilled by an X-Pro-Gly tripeptide. The hydroxylation reaction requires Fe^{2+}, 2-oxoglutarate, O_2 and ascorbate. The 2- oxoglutarate is stochiometrically decarboxylated during hydroxylation, with one atom of O_2 being incorporated into succinate and the other into the hydroxy group formed on the proline residue [*Kivirikko and Pihlajaniemi,* 1998].

Kinetic studies have concluded that there is a sequential binding of Fe^{2+}, 2-oxoglutarate, O_2 and the peptide substrate to the enzyme in this order, and also an ordered release of the hydroxylated peptide, CO_2, succinate and Fe^{2+}. Oxygen is probably activated as superoxide. Ascorbate is not consumed stoichiometrically, and P4H can catalyze a number of reactions in its absence. However, when the peptide substrate is present in a saturating concentration, P4H can decarboxylate without hydroxylation of proline and the ascorbate acts as an alternate O_2 acceptor [*Kivirikko,* 1980*; Kivirikko and Pihlajaniemi,* 1998].

The action of prolyl 4-hydroxylase is the key for achieving stability of the procollagen molecule under physiological conditions. Without an appropriate number of hydroxylated Y-position prolyl residues, the newly synthesized chains cannot efficiently fold into a triple helical conformation at 37°C[*Kivirikko and Pihlajaniemi,* 1998].

Moreover, if insufficient hydroxylation occurs, the polypeptides will remain non-

helical, resulting in their degradation, poor secretion and ineffective self-assembly into collagen fibrils [*Pihlajaniemi et al.,* 1987*; Vuorela et al.,* 1997*; Kivirikko and Pihlajaniemi,* 1998].

Prolyl 3-hydroxylase, also belongs to the group of 2-oxoglutarate dioxygenases and, like prolyl 4-hydroxylase, requires Fe^{2+}, 2-oxoglutarate, O_2, and ascorbate for its activity [*Kivirikko and Pihlajaniemi,* 1998*; Vranka et al.,* 2004]. 3-hydroxyproline is found in almost all collagens in the sequence Gly-Pro-Pro-Gly, with the hydroxyl group in the first proline of the tripeptide. The largest amounts of 3-hydroxyproline are found associated with collagen types IV and V, but the occurrence of 3-hydroxyproline is much less frequent than that of 4-hydroxyproline in the total amino acid content of collagens. In basement membrane collagens where 3-hydroxyproline has been reported to be most abundant, the total content is around 115 residues per 1000 residues, which is just 10% of that of 4-hydroxyproline [*Vranka et al.,* 2004].

As stated above, 4-hydroxyproline, has been shown to be important for the stabilization of the triple helix. In contrast, the role of 3-hydroxyproline in collagens is not well understood. Recent studies conducted on embryonic chick cells showed that P3H1 can be purified from the rough endoplasmic reticulum, and also that the enzyme is present in a complex of proteins that specifically bind to denatured collagen. Since the study used a denatured fibrillar collagen as the affinity substrate, and because the prolyl 3-hydroxylase 1 immunolocalization correlates it with the presence of fibrillar collagens, it is more likely that the enzyme modifies fibrillar collagens; it also supports

the idea that P3H plays an important biological role in the folding and assembly of triple helical collagen [*Vranka et al.,* 2004].

Lysyl hydroxylase is an $\alpha 2$ homo-dimer of about 200,000 kDa, that hydroxylates the X-Lys-Gly repeated sequence by a mechanism similar to P4Hs. In human, three isoforms have been isolated (LH1, LH2 and LH3) that have lysyl hydroxylase activity. The number of hydroxylated lysyl residues varies among different collagen types and tissues, depending on the physiological condition. The function of these hydroxylysyl groups is to serve as substrates for glycosylation before the chain becomes helical, and they are essential for cross-linking of collagen chains [*Kivirikko and Pihlajaniemi,* 1998*; Kivirikko,* 1980].

Glycosylation of hydroxy-lysyl residues is catalyzed by two enzymes, namely hydroxylysyl galactosyltransferase that transfers galactose to hydroxylysyl residues and galactosylhydroxylysyl glucosyltransferase, that transfers glucose to galactosylhydroxylysyl residues [*Kivirikko,* 1980]. Studies conducted with insect cells and *E. coli* revealed that glycosylation of some hydroxy-lysyl residues to galactosylhydroxylysine and glucosylgalactosyl-hydroxylysine is catalyzed also by LH3 [*Wang et al.,* 2002].

Asparagine residues in the C-terminal propeptide are likewise glycosylated. After C-propeptides are associated, disulfide bonds are formed and, if an average of 100 4-hydroxyproline residues have been formed on each of the three α chains, a nucleus is formed in the C-terminal region and chain formation is propagated in a zipper-like fashion [*Engel and Prockop,* 1991].

Translocation and secretion of procollagen:

The procollagen molecule is transported from the endoplasmic reticulum across the Golgi stacks without leaving the lumen of the Golgi cisternae. During this transfer process, the molecules begin to aggregate laterally, resulting in condensation and formation of granules that will be secreted to the extracellular space [*Prockop and Kivirikko,* 1995*; van der Rest and Garrone,* 1991].

Extracellular modifications:

Extracellular modifications are enzymatically performed by specific proteinases that cleave the C- and N-propeptides. Both proteinases require calcium as a bivalent cation and, as a result the procollagen molecules are converted into insoluble collagen [*McLaughlin and Bulleid,* 1998*; Myllyharju and Kivirikko,* 2001].

The collagen molecules aggregate spontaneously into fibrils. Cross-link formation is the last step in the process of collagen synthesis. It involves the oxidative deamination of ϵ-amino groups in lysyl and hydroxylysyl residues and is catalyzed by lysyl oxidase. The aldehydes generated in the lysyl oxidase reaction can serve either for cross link formation or intramolecular cross-links [*Kadler et al.,* 1987*; Kadler et al.,* 1996*; Prockop et al.,* 1998].

1.3 Scope and Objectives

The long term goal of this research is to produce procollagen I in transgenic barley grains. Several recombinant human and non-human proteins have been expressed in the barley grain, including lysozyme, lactoferrin [*van Fleet*, 2001] and an engineered heat stable β-glucanase [*Horvath et al.*, 2000]. The aim of the present research is to synthesize components needed for the production of procollagen in the endosperm of transgenic barley.

In order to achieve this goal, this study focuses on two aspects of the production of recombinant collagen in barley. First, a gene codon-optimized for plant expression had to be synthesized and its functionality tested in an eukaryotic expression system. Towards this goal the COL1A1 gene was synthesized to yield a gene that matched the monocot plants codon usage, with a GC content over 60%, and inserted into a vector for expression in *Pichia pastoris*. Likewise a vector carrying the genes for prolyl 4-hydroxylase for synthesis of hydroxylated procollagen in the yeast *Pichia pastoris* had to be constructed and expressed. The second goal was to obtain barley plants containing the genes for procollagen synthesis. Plasmid vectors containing the genes for the assembly of human type I homotrimeric procollagen had to be prepared for *Agrobacterium tumefaciens* mediated transformation in barley. The objectives of the study can be summarized as follows:

- Synthesis of plant codon-optimized human gene encoding for α1(I) chain (COL1A1), assembly of the vectors needed for protein expression and expression of correctly

assembled homotrimeric human type I procollagen in *Pichia pastoris*.

- Development of vectors needed for transformation in barley and generation of barley plants containing procollagen I and prolyl 4-hydroxylase genes by *Agrobacterium tumefaciens*-mediated transformation.

1.4 Thesis Outline

This dissertation is organized into three main chapters. In Chapter 2 I discuss assembly of the gene, development of vectors and the expression of homotrimeric procollagen I and prolyl 4-hydroxylase in *Pichia pastoris* used as test for the correctness of the construct.

In Chapter 3 I present the results obtained in the development of the vectors needed for the *Agrobacterium*-mediated transformation and preliminary results obtained from the co-cultivation of immature zygotic barley embryos.

Tables and figures are presented at the end of each chapter.

1.5 Tables and Figures

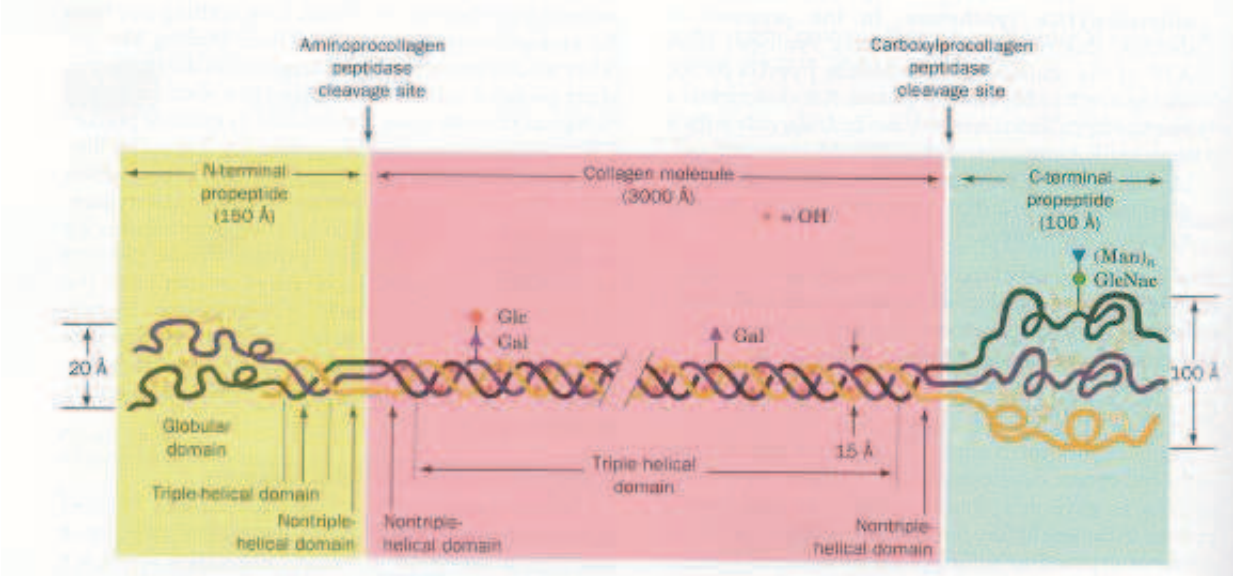

Figure 1.1: Collagen model

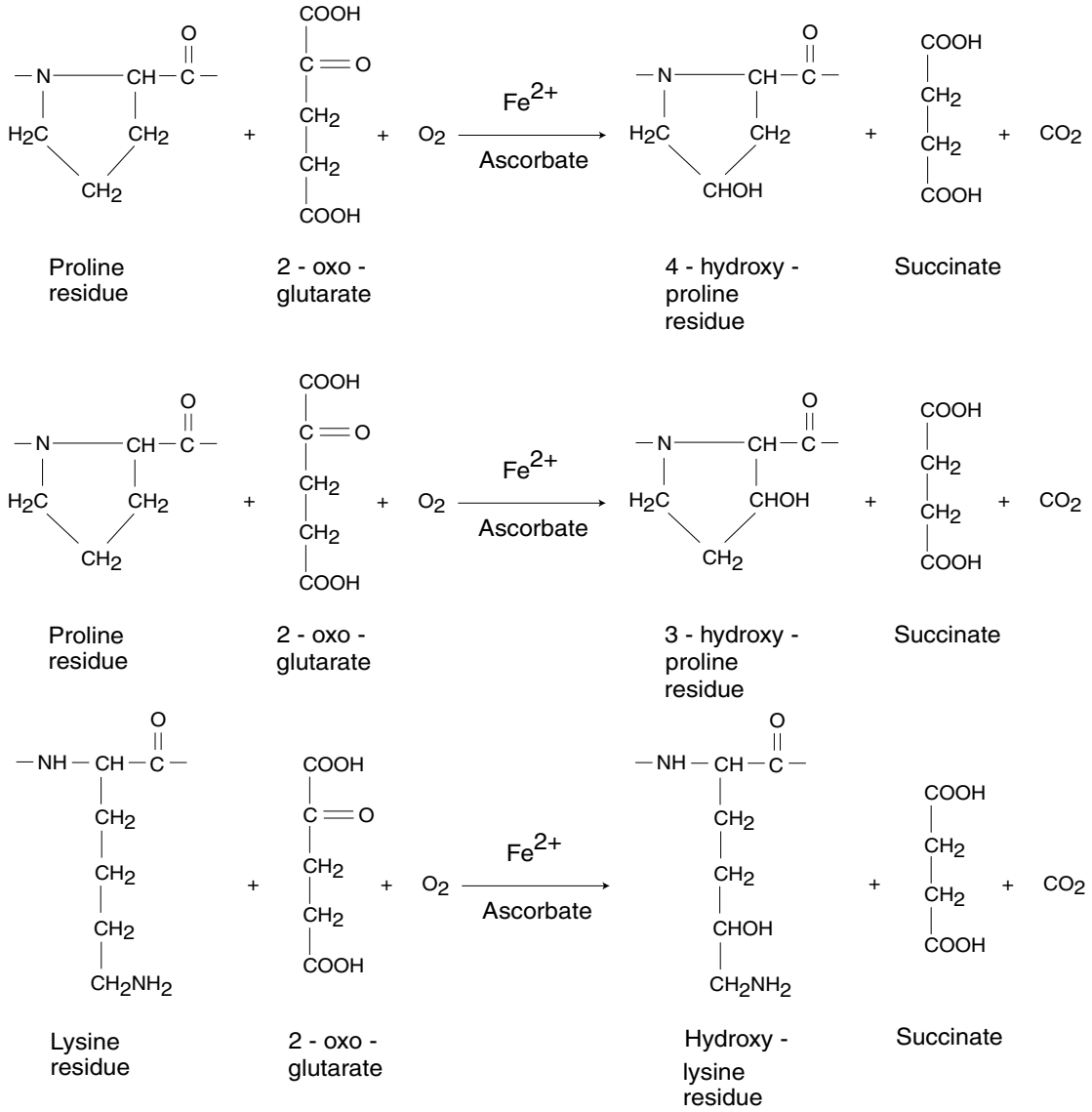

Figure 1.2: Reactions catalyzed by prolyl 4-hydroxylase, prolyl 3-hydroxylase and lysyl hydroxylase [*Kivirikko and Pihlajaniemi,* 1998]

Table 1.1: Collagen types, occurrence and polymeric structure.

Polymeric structure	Collagen type	Occurrence
Fibril-forming collagens	I, II, III, V and XI	Connective tissues, cartilage, vitreous humor, extensible connective tissue
FACIT and related collagens	IX, XII, XIV, XVI, XIX	Associated with collagen I and II
Collagen forming hexagonal networks	VIII and X	Endothelium and hypertrophic cartilage
Family of type IV	IV	Basement membranes
Collagen forming beaded filaments	VI	Most connective tissues
Collagen forming anchoring fibrils	VII	Anchoring fibrils
Collagen with transmembrane domains	XIII and XVII	Skin hemidesmosomes and many tissues
Family of type XV and XVIII	XV and XVIII	Many tissues, specially liver and kidney

[*Myllyharju and Kivirikko,* 2001*; Prockop and Kivirikko,* 1995*; Bateman et al.,* 1996]

Chapter 2

Collagen I expression in *Pichia pastoris*

2.1 Abstract

Collagen I is a main constituent of the extracellular matrix composed by three chains, two α1(I) chains and one α2(I) chain that assembles into a triple helix with a coiled structure. For collagen to be correctly expressed, eight specific enzymes are required, among them, prolyl 4-hydroxylase plays a central role, since it catalyzes hydroxylation of prolyl residues that confer thermal stability to the newly synthesized chain. The main objective of this study was to assemble the codon-optimized gene that encodes the (α)1 chain of collagen type I and express this gene in *Pichia pastoris*. The following specific objectives were addressed: 1. synthesize the gene coding for the α chain of human collagen type I (COL1A1) with codons optimized for translation on barley polysomes in the endosperm; 2. develop the vectors needed for protein expression in *Pichia pastoris*; and 3. synthesize COL1A1 and prolyl-hydroxylase in *Pichia pastoris*. Using the degeneracy of the genetic code, a collagen gene was designed *in silico* with

nucleotides in the third position of the codons to reach a GC content of 67% as required for optimal translation in the barley endosperm. The gene was assembled with intermediate vectors containing 64 overlapping oligonucleotides that covered both strands of the DNA double helix. Synthesis and secretion from *Pichia pastoris* was obtained with vectors containing the gene under the control of an alcohol oxidase promoter and insertion of the *Saccharomyces cerevisiae* α mating type secretion signal for export of the protein into the medium. SDS-PAGE and Western blotting with specific antibodies recognizing 25kDa gelatin identified a protein present in the cell lysate and culture medium as unhydroxylated homotrimeric procollagen I. An ELISA test was conducted in order to quantify the amount of protein produced.

2.2 Introduction

Collagen I is a polypeptide composed of three chains: two α chains and one β chain that assemble into a triple helix with a coiled structure[*Prockop et al.,* 1998*; Kivirikko,* 1980*; van der Rest and Garrone,* 1991].

Synthesis of collagen involves an unusual number of post-translational modifications, that require at least eight specific enzymes[*Fessler and Fessler,* 1978*; van der Rest and Garrone,* 1991*; Myllyharju,* 2003]. Among these enzymes, prolyl 4-hydroxylase (P4H), which in vertebrates is an $\alpha 2\beta 2$ tetramer, has a central role: it forms the 4-hydroxyproline residues required for folding of the newly synthesized collagen polypeptide chains into triple-helical molecules and thus procures the thermal stability needed

for living organisms.

Recombinant collagen has been synthesized in several yeast species, and increasing efforts are being made to develop efficient and safe alternatives to animal-derived collagens. The main advantage of yeasts compared to bacteria lies in their capacity to perform post-translational modifications and efficient secretion of eucaryotic proteins. Yeasts were considered to lack prolyl-hydroxylases; this is the case for *Pichia pastoris* and *Saccharomyces cerevisiae* [*de Bruin et al.,* 2002].

Pichia pastoris [*Vuorela et al.,* 1997] and *Saccharomyces cerevisiae* [*Toman et al.,* 2000] are able to express hydroxylate and correctly fold the triple helical collagen polypeptide only with transgenic co-expression of the hydroxylation enzymes. However *Hansenula polymorpha* [*de Bruin et al.,* 2000] synthesizes 4-hydroxyproline and is thus able to form prolyl 4-hydroxylated collagenous domains.

In the case of *Saccharomyces cerevisiae*, expression of α1(I) procollagen without prolyl hydroxylase genes resulted in low levels of expression, but the system was able to generate the triple helical procollagen, thereby opening the possibility that endogenous yeast PDI can assemble the three procollagen polypeptide chains, but in the absence of α subunit of prolyl 4-hydroxylase, the generated procollagen lacked thermal stability (Tm of 23- 25°C), and was degraded by pepsin. When collagen was expressed with the genes encoding prolyl hydroxylase, the triple helical molecules produced were thermally stable with a Tm 35°C[*Toman et al.,* 2000] and remained inside the endoplasmic reticulum. This experiment led to the conclusion that the genes needed for the efficient

assembly of type I collagen are the ones that encode COL1A1 and COL1A2 chains and both subunits of prolyl 4-hydroxylase. Additionally, for higher levels of expression, glutamate was also required as a precursor for the synthesis of α-ketoglutarate.

On the other hand, studies conducted with *Pichia pastoris* revealed that production of stable prolyl 4-hydroxylase tetramer requires expression of collagen type III chains, and at the same time, tetramer assembly is a requisite for stable triple helical collagen formation [*Vuorela et al.,* 1997]. Nevertheless, only 10% of the molecules assembled in *Pichia pastoris* were secreted into the medium [*Keizer-Gunnink et al.,* 2000], an observation that led to the conclusion that the chains are retained within the endoplasmic reticulum, probably due to the large size of the procollagen molecules.

Myllyharju and collaborators, originally regarded the C-propeptides of procollagens as essential for correct chain recognition and assembly [*Myllyharju and Kivirikko,* 2001], but it was subsequently found that the C-propeptides can be replaced in the *Pichia pastoris* expression system by foldon, a 29-residue trimerization domain located at the C-terminus of the bacteriophage T4 fibritin molecule [*Pakkanen et al.,* 2003]. Foldon turned out to be more effective in forming procollagen trimers than the C-propeptides. Furthermore, co-expression of α1(I)-foldon and α2(I)-foldon chains led to effective assembly of heterotrimeric molecules with the expected 2:1 chain ratio. As foldon contains no information for chain recognition, the data indicate that chain selection and assembly of procollagen are influenced not only by the C-terminal oligomerization domain but also by determinants present in the central part of the

collagen chains.

Comparing the expression and secretion of a procollagen I fragment in *Hansenula polymorpha* and *Pichia pastoris* led to the conclusion that *H. polymorpha* can secrete a recombinant human fragment of procollagen I [*de Bruin et al.,* 2000]. In a later study [*de Bruin et al.,* 2002], procollagen I produced was hydroxylated by an internal mechanism triggered by peptone present in the culture media.

The aim of this dissertation is to produce transgenic plants expressing the homotrimeric procollagen I and prolyl 4-hydroxylase genes. Since it has been shown that monocot plants prefer different codons than mammals to express the same amino acid, the $\alpha(1)$ collagen type I gene was optimized in order to ensure an adequate usage of the protein code [*Horvath et al.,* 2000*; Wu,* 2003*; van Fleet,* 2001*; Horvath et al.,* 2001]. A collagen gene with a GC content of 67% was synthesized, while the natural GC content of the structural genes for the prolyl 4-hydroxylase genes coding for the α subunit (63%) and β subunit (60%) was considered adequate.

The goal of the work described in this chapter was to express codon optimized procollagen type I in the methylotrophic yeast *Pichia pastoris* in order to verify the correct structure of the synthesized gene.

Specifically, the following objectives were addressed:

- Synthesis of $\alpha 1$ chain of human collagen type I gene with codons optimized for translation on barley polysomes in the endosperm.

- Development of the vectors needed for protein expression in *Pichia pastoris.*

- Synthesis of COL1A1 and prolyl 4-hydroxylase in *Pichia pastoris*

2.3 Materials and Methods

Strains

Escherichia coli strain DH5α: Plasmids were cloned in *E. coli* strain DH5 (α sup E44Δlac U169 (Φ80 lacZΔM15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1).

Yeast strains: Recombinant proteins were expressed in strains GS115 and X-33 of the methylotrophic yeast *Pichia pastoris*. GS115 has a mutation in the histidinol dehydrogenase gene (*his4*) and is therefore autotrophic for histidine synthesis [*Cregg et al.*, 1985]. Expression plasmids carrying the HIS4 gene can complement *his4* in the host strain, and the transformants can be selected for their ability to grow on histidine-deficient medium. X-33 is a wild type *Pichia* strain that is useful for selection on Zeocin® and for large-scale growth. Expression plasmids carrying the *Sh ble* gene confer resistance to Zeocin® to both strains.

Plasmids

pUC18 (see Fig. 2.5) plasmid from MBI Fermentas (Hannover, MD) was used as an intermediate cloning vector. It is a high copy-number plasmid that has the pMB1 replicon responsible for the replication of the plasmid source and the *bla* gene that confers resistance to ampicillin.

pICZA (Invitrogen, Carlsbad CA)was used as an intermediate cloning vector. It

has the *Sh ble* gene for selection on Zeocin® containing media.

pICZAαA from Invitrogen (see Fig. 2.6) provides the *Saccharomyces cerevisiae* α mating factor signal peptide to the N terminal, and the histidine tag to the C-terminal end of the targeted recombinant protein for secretion and purification respectively. The plasmid also carries the 5'AOX1 gene promoter that allows methanol-inducible, high-level expression of the target gene in *Pichia.* It has *Sh ble* gene for selection in Zeocin® containing media.

pHIL-S1 plasmid from Invitrogen was employed for expression in *Pichia pastoris.* It has the HIS4 gene for selection of recombinant strains able to grow on histidine-deficient medium. It also has the AOX1 promoter that drives the transcription of the target gene and PHO1 secretion signal for secretion of the protein into the culture medium. It has the ampicillin resistance gene and an *E.coli* origin of replication for maintenance, selection and replication of the vector in *E. coli* (see Fig. 2.7).

Growth media: *E. coli* cells were cultivated in liquid LB media or on LB plates according to protocols of Sambrook et al 1989 [*Sambrook et al.,* 1989]. *Pichia pastoris* was cultivated according to protocols by Invitrogen (Carlsbad, CA).

Enzymes used were from MBI Fermentas (Hannover, MD) if not otherwise stated.

Synthesis of the gene coding for codon-optimized $\alpha(1)$ chain of human collagen type I

The gene coding for codon optimized $\alpha(1)$ chain of human collagen type I (NCBI accession number 000088) was codon-optimized in order to match the codon usage for monocot plants with a GC content over 60%. Sixty seven oligonucleotides were synthesized for assembly of the double strand of the collagen 1 codon-optimized gene containing the N- and C-telopeptides and foldon (IDT, Coralville, IA). Oligos 1-33 are orientated in the 5' direction and oligos 34R-67R are orientated in the 3' direction. Each oligo was 100 base-pairs long with a 50 base-pair overlapping region with the complementary oligo on the opposite strand. Oligos 1 and 64 have five base pairs of the D-hordein signal peptide coding sequence for easy cloning into barley expression vectors (for a detailed list of oligos see Appendix 1).

The oligos were separated into four groups. Two hundredth of a microliter of each oligo was added to an Eppendorf tube and the volume adjusted up to 50μl with sterile water. For each group two reactions were carried out with 1 and 0.5μl aliquots, respectively, in order to safeguard for optimal concentration. The reactions were incubated in boiling water for 3 min and cooled overnight to allow complete annealing of the oligos. The annealed oligos were ligated with *Pfu* ligase (Promega, Madison, WI) at 55°Cfor 1 h, and 1μl of the reaction mixture was used as template for polymerase chain reaction (PCR) with primers described in table 2.1 using program COL2(25) (see Appendix, Table 5.2).

The PCR product was subjected to agarose gel electrophoresis (1% gel, 120V, 1h), and a single band of the expected size was cut out, purified with Qiagen gel extraction kit and phosphorylated with T4 polynucleotide kinase according to the manufacturer's protocol.

Four fragments were cloned into *Sma*I digested and dephosphorylated pUC18 according to standard protocol [*Sambrook et al.,* 1989](see Figs. 2.8, 2.9, 2.10 and 2.11). The clones were sent to Amplicon Express (Pullman, WA) for sequencing. One mutation was found in clone pUC18(5'sp-3'2), and it was corrected by splicing by overlapping extension PCR (SOE-PCR) using the protocol described by Horvath, 2000 [*Horvath et al.*, 2000].

For the assembly of the gene, four unique restriction sites were used in order to generate two intermediate vectors. In the first case, pUC18(5'sp-3'2) was digested with *Hin*dIII/ *Eag*I, the fragment purified and cloned into pUC18(5'2-3'5) that had been digested with the same enzymes. The plasmid generated was named pUC18(5'sp-3'5)(see Fig. 2.12).

pUC18(5'5-3'8) was digested with *Hin*dIII/ *Nhe*I and ligated into pUC18(5'8-3'10) previously digested with the same enzymes to yield plasmid pUC18(5'5-3'10)(see Fig. 2.13).

A three-way ligation was performed in order to assemble the 3261 base pairs corresponding to the optimized gene. pICZA was used as a donor vector; it was digested with *Hin*dIII and *Eco*RI. On the other hand, two fragments corresponding to half of

COL1A1 each were digested from plasmid pUC18(5'sp-3'5) and pUC18(5'5-3'10) with *Hin*dIII/ *Bgl*II and *Eco*RI/ *Bgl*II respectively. The three pieces were ligated according to manufactures protocol and colonies grown on Zeocin® containing media were isolated and DNA was extracted using Bio-Rad(Hercules, CA) mini-prep kit. Correct orientation of the insert was checked by restriction site analysis to yield plasmid pICZA(COL1A1).

2.3.1 *Pichia pastoris* vector construction

In order to generate the vectors suitable for expression in *Pichia pastoris*, it was necessary to eliminate the five base pairs at the beginning of the gene, delete the stop codon at the end, and add *Eco*RI and *Not*I restriction sites at the beginning and at the end of the collagen 1 gene. This step was carried out by modifying the beginning and the end of the gene with a polymerase chain reaction (PCR).

Primers used were EcoRI/ 3'2 and 5'8/ NotI. Two fragments were amplified from pUC18(5'sp-3'2) and pUC18(5'8-3'10), cloned into pUC18 and sequenced (Amplicon Express, Pullman, WA). The fragments were then digested with *Hin*dIII and *Eco*RI, respectively, blunted and digested with *Xba*I and *Nhe*I and, after purification, the corresponding segments were placed into pICZA(COL1A1) to generate pICZA(EcoRI-proCOL1A1-NotI). The vector was transformed into DH5α and transformants resistant to Zeocin® were selected. DNA was isolated and digested with *Eco*RI/*Not*I and ligated into pICZAαA to generate pICZAαA(EcoRI-proCOL1A1-NotI)(see Fig. 2.14)

The gene of the α subunit of the prolyl 4-hydroxylase was digested from plasmid pPCRScript-H4Palfa (Fibrogen, San Francisco, CA) with *Eco*RI and *Not*I and ligated into the multiple clonning site of pHIL-S1 to generate pHIL-S1(α)as shown on Fig. 2.19.

The gene encoding β subunit of the prolyl 4-hydroxylase was digested from plasmid pBlue-P4HmApt (Fibrogen, San Francisco, CA) with *Eco*RI and *Not*I and ligated into *Eco*RI/*Not*I digested pICZαA to yield plasmid pICZαA(β)(see Fig. 2.16).

2.3.2 Transformation, growth and induction of *Pichia pastoris* strains

Strain GS115 was used as a parental strain of *Pichia pastoris* (for a detailed description, see p. 23). Transformation was carried out by the easy comp method (Invitrogen,Carlsbad, CA),which is based on a protocol described by [*Cregg et al.,* 1985]. The procedure is carried out by a chemical treatment of the yeast cells with a sorbitol solution that contains ethylene glycol and DMSO in order to make the cells competent. This is followed by the addition of a PEG solution for the transformation itself. pICZAα(EcoRI-proCOLA1-NotI) was linearized with *Sal*I and transformed into the host strain by the above method. Transformants were selected on YPD(+Zeocin$^{\textcircled{R}}$) plates.

pHIL-S1(α)was linearized with *Sac*I and transformed into GS115 following the manufacturers protocol (Invitrogen, Carlsbad, CA). Transformants were selected by

their ability to grow on minimal media plates without histidine.

pICZαA(β)was linearized with *Sal*I and transformed into the host strain GS115. The transformants were selected on YPD +Zeocin® plates.

A fourth strain was generated to express both α and β subunits of prolyl 4-hydroxylase. GS115 was transformed following the above protocol with pHIL-S1(α) *Sac*I linearized and pICZαA(β) *Sal*I linearized in order to generate a strain that produces the α2-β2 tetramer. Transformants were selected on minimal plates without histidine and Zeocin®.

The cells were cultured according to the manufacturers protocol with modifications as described in Vuorela et al, 1997[*Vuorela et al.*, 1997]. Cells were cultured in 250 ml shaker flasks in a buffered glycerol complex medium, pH 6.0 with 10g/liter of yeast extract and 20g/liter peptone. Expression was induced in complex buffered methanol medium, pH 6.0 and methanol was added every 24 hours to a final concentration of 1%. Amino acids were added as required. One milliliter aliquots were harvested every 24 h, up to 96 h. Supernatant was concentrated with Centricon YM50 (Millipore, MA), and the cells were resuspended in cold breaking cell buffer and broken by vortexing with acid-washed glass beads, and the lysate then centrifuged for 10 minutes at maximum speed at room temperature. Aliquots of the concentrated liquid culture and soluble fractions were analyzed by 7.5% SDS-PAGE under non-reducing conditions and the gel stained with Coomassie blue. The gel was transferred to a PVDF membrane using a semi-dry method. Eighteen volts were applied for 1 hour before blotting the membrane

with 2% skim milk. After 1 h of blotting, the membrane was decorated with specific antibodies for collagen (primary antibodie 50kDa pooled rabbit anti-sera, Fibrogen; secondary antirabbit IgG whole molecule peroxidase conjugate, Sigma) and stained for visualization.

2.3.3 Direct polymerase chain reaction (PCR) screening

In order to check the integration of the expression plasmid into the *Pichia* host genome, Zeocin® resistant colonies were screened by PCR based on the protocol by Linder et al [*Linder et al.,* 1996]. Yeast cells were lysed by a combination of lyticase, freezing and heating. A sample of genomic *Pichia* DNA was taken and the crude lysate subjected to PCR with primers and program shown in the Appendix 5, Table 5.3.

2.3.4 ELISA test

One hundred microliters of culture medium and cell lysates were used for coating the wells of a Immulon 2HB plate (Dynatech, Alexandria, BC). The pH of the culture medium was increased by addition of 200mM sodium bicarbonate and the medium kept 4 hours at room temperature for binding. The medium was discarded and the wells were washed three times with TBS-Tween and incubated with 200μ of blocking buffer for 2 hours to block unreactive groups. The wells were washed again with TBS-Tween and incubated at 4°Covernight with primary antibodies diluted 1:500 in 2% skim milk. After washing, secondary antibody was added and incubated for 4 hours

at room temperature. Then the wells were stained with a solution containing 2ml Tris pH 8.0, 100μl 3% H_2O_2 and 1ml of 4-chloronaphtol (50mg/ml) in 100ml of water. Pichia derived collagen and bovine gelatin were used as controls at concentrations of 100, 10 and 2 ng/ml.

2.4 Results and Discussion

2.4.1 Synthesis of COL1A1

The gene was codon optimized in order to match the monocot plant codon usage for amino acids. As stated by Batard and collaborators [*Batard et al.,* 2000], synonymous codons are not equally utilized in all organisms. Within different species, the codon bias becomes stronger for highly expressed genes, probably due to the need to ensure accuracy and efficiency of translation, and to match abundance of the corresponding tRNAs [*Chiapello et al.,* 1998]. A change in codon usage also implies changes in GC content, in secondary structures, in translation initiation and codon context, and in potential splice sites of mRNA. Alone or together, these factors affect the efficiency of translation, often resulting in proteins being poorly expressed or erroneously translated in heterologous organisms [*Batard et al.,* 2000*; Koziel et al.,* 1996]. In order to avoid undesired effects due to expression of a mammalian protein in barley, the gene encoding for collagen 1 was codon optimized according to monocot codon usage [*Murray et al.,* 1989]. The optimized sequence is shown in figures 2.1, 2.2, 2.3 and 2.4. The N-

and C-terminal sequences are shown in red and the central part corresponding to the α1 chain is shown in black. The final portion corresponding to the foldon sequence is shown in blue. The final GC content was 67%.

The gene encoding for the α1 chain of collagen I was synthesized starting with 64 oligos that were assembled as described in Methods. Using intermediate vectors, 3261 base pairs corresponding to the gene were finally assembled as shown in Fig. 2.17. The gene was sequenced at all intermediate steps; one mutation was corrected by Splicing by Overlap Extension Polymerase Chain Reaction (SOE-PCR) as described by Horvath et al [*Horvath et al.*, 2000].

2.4.2 Construction of expression vectors and transformation of *Pichia pastoris* strains

Three expression vectors were constructed for the expression of codon-optimized homotrimeric procollagen type I and α and β subunits of prolyl 4-hydroxylase. The first plasmid contained the codon-optimized gene coding for α1 chain of collagen I, the second contained the α subunit of prolyl 4-hydroxylase and the third vector contained the β subunit of prolyl 4-hydroxylase.

The expression plasmids were transformed into the *E. coli* strain DH5α, and the clones subjected to restriction enzyme analysis to verify integration of the fragments with sizes of 3261, 1621 and 1489 base pairs, respectively. Restriction digest of the plasmid with the collagen gene is shown on Fig. 2.18. The plasmid was linearized with

*Eco*RI and cut with two different restriction enzymes giving the sizes expected. The digest of the plasmid containing the α subunit of prolyl 4-hydroxylase was digested with *Hin*dIII that cuts the gene further, specifically in the AOX promoter and terminator. The figure is presented in Fig. 2.19. The last expression plasmid is that for the β subunit of prolyl 4-hydroxylase (2.20) that was digested with *Eco*RI and *Not*I that cut at the ends of the gene.

Plasmid DNA was isolated and *Pichia pastoris* strain GS115 was transformed following the EasyComp method (Invitrogen, CA). The genes were inserted into a multiple cloning site downstream of the strong *P. pastoris* AOX1 alcohol oxidase promoter, to effect transcription induction by methanol. In the case of the P4H α subunit, the gene was ligated with the code for its own signal peptide. On the other hand, COL1A1 and the P4H β subunits genes were ligated in frame with the pre-pro α-mating factor sequence of *S. cerevisiae* to facilitate secretion of the peptides. The pic$Z\alpha$A plasmid confers resistance to Zeocin®, whereas, pHIL-S1 will allow the yeast strain to grow on minimal medium. Both plasmids lack a yeast origin of replication; therefore, Zeocin® resistant and His+ colonies generated after transformation are attributable to integration of the vector(s) into the yeast genome. A summary of the vectors generated, the transformed strains and the polypeptide encoded is presented in table 2.2. All transformed colonies used for protein expression analysis were scored for their phenotype, meaning that the integration of the plasmid was at the 5' region of the AOX1 gene. As a result, the transformants grow normally with methanol as the sole carbon source.

Additionally, colonies were screened by direct PCR to confirm the integration of the plasmid into the yeast genome (Fig. 2.21, 2.22 and 2.23).

2.4.3 Protein expression

Pichia pastoris strain GS115 used as host strain; it is auxotroph for histidine and lacks the *Sh Ble* gene that confers resistance to Zeocin®.

In order to evaluate the correct function of the codon modified collagen type I gene, the gene was inserted into *Pichia pastoris* expression vector pICZAαA, and the host strain GS115 then transformed to generate GS115(pICZAαA(EcoRI-proCOL1A1-NotI)). The strain was methanol-induced, and cells and supernatant were collected after 96 hours. Cells were broken with glass beads in an appropriate phosphate buffer, and the supernatant was concentrated using YM50. Both supernatant and cell lysate were analyzed without pepsin treatment by 7.5% SDS-PAGE followed by Coomassie blue staining or with Western blots using specific antibodies against the collagen type I domain for visualization (Fibrogen, San Francisco, CA)

A band corresponding to the size of homotrimeric procollagen I was observed on the SDS-PAGE gel upon Coomassie staining (Fig. 2.24) from the cell lysate and concentrated culture media. The protein band was confirmed to be procollagen by reacting in a Western blot with antibodies specific for an unhydroxylated procollagen1 domain (Fig. 2.25). The size of the protein was about 180 kDa, which is in the range of non hydroxylated homotrimer procollagen I according to Myllyharju and collabora-

tors [*Myllyharju et al.,* 1997], who expressed collagen type I with bacillovirus in High Five insect cells. Vuorela at al [*Vuorela et al.,* 1997] obtained homotrimeric collagen I in *Pichia pastoris*, with the same apparent molecular size as in the present study. In a recent report, the same group, [*Pakkanen et al.,* 2003], successfully expressed procollagen I plus foldon in *Pichia pastoris* and the apparent size was also about 180kDa . The level of expression, nevertheless, was very low, requiring concentration of the supernatant with Centricon filters in order to visualize the protein by Coomassie staining in SDS-PAGE gels or as Western blots. This could have been due to either the high molecular weight or to inefficient transport through the yeast cell wall. Previous analysis [*Myllyharju et al.,* 2000*; Nokelainen et al.,* 2001], in experiments with collagen I and III concluded that the protein accumulates inside the endoplasmic reticulum of the yeast cell [*Keizer-Gunnink et al.,* 2000]. Surprisingly, no significant collagen could be detected in our study inside the yeast cell. One possible explanation could be that the collagen was degraded inside the cell or that the protease inhibitor (phenylmethylsulfonyl fluoride) added to the breaking buffer did not act as expected. A second explanation for the low level of expression could be an insufficient amount of O_2 provided in the shaker flask during induction of protein production [*Myllyharju et al.,* 2000]. A third factor that could determine the level of expression was the use of the α mating factor signal peptide, which also has given low yield in a previous study [*Vuorela et al.,* 1997]. Finally it is possible that the collagen molecule aggregates with endogenous proteins, giving a higher apparent molecular size and a non-specific smear

with the Western blot

Expression of prolyl 4-hydroxylase β subunit was achieved by cultivating the strain in buffered media. After 72 hours, cells were broken as described above and the supernatant was concentrated with Centricon YM30. Both supernatant and the soluble fraction from the cell lysate were analyzed by 10% SDS-PAGE followed by Coomasie staining of the gel. A 55kDa protein band corresponding to the apparent molecular size was visualized in the Coomassie stained SDS-PAGE (Fig. 2.26). The expression of this protein is in agreement with the results of Vuorela et al [*Vuorela et al.,* 1997] who stated also that expression of PDI with the yeast prepro α-mating factor sequence increased the yield of β subunit of P4H secreted into the medium compared to the construct with its own signal peptide or that with the *Pichia pastoris* acid phosphatase 1 signal sequence.

In the case of the expression of α subunit of P4H, no detectable level of expression was achieved by induction of the yeast cell. On the other hand, when the strain containing both genes for the expression $\alpha 2 \beta 2$ tetramer was induced, only a band corresponding to β subunit was visualized on the gel (see Fig. 2.27). This can be explained by instability of the tetramer in the absence of procollagen molecules; consequently, the α subunit is degraded and only the β subunit remains [*Vuorela et al.,* 1997].

2.5 Tables and Figures

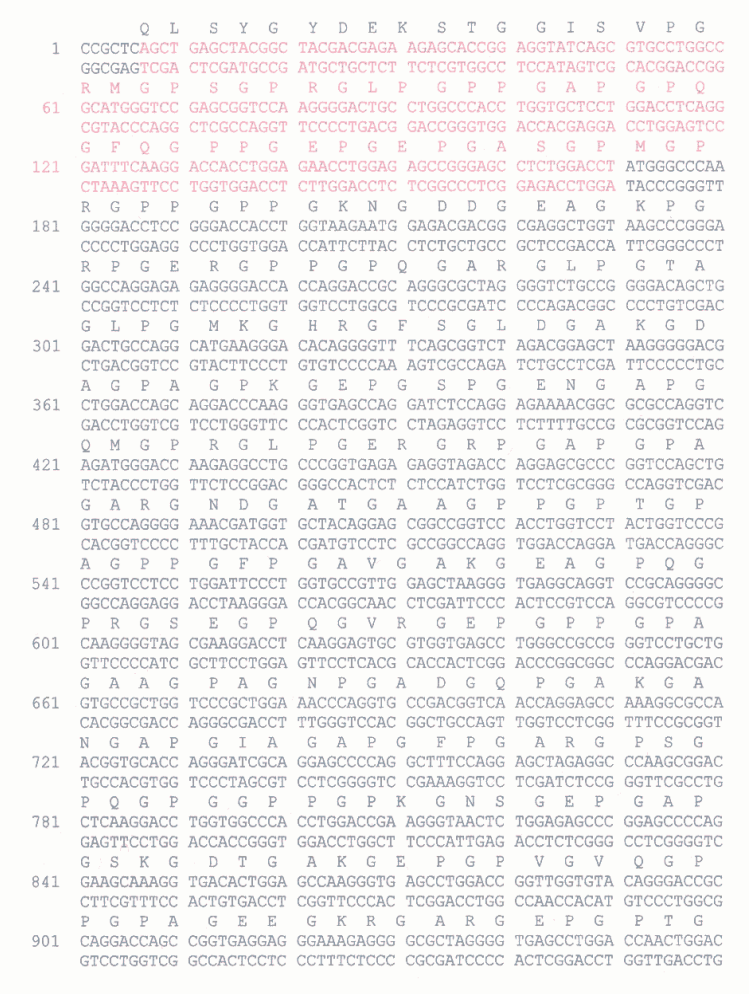

Figure 2.1: N- and C-telopeptide regions (red). The collagen domain (black) and foldon (blue)of COL1A1 sequence. Part 1

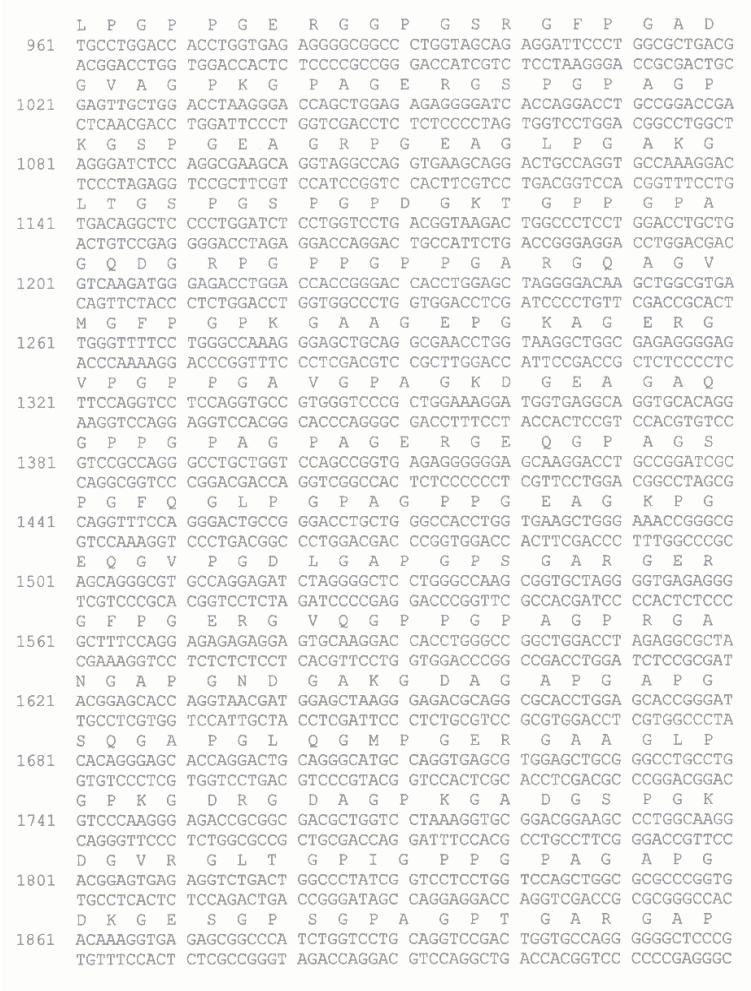

Figure 2.2: Codon optimized COL1A1 sequence, part 2

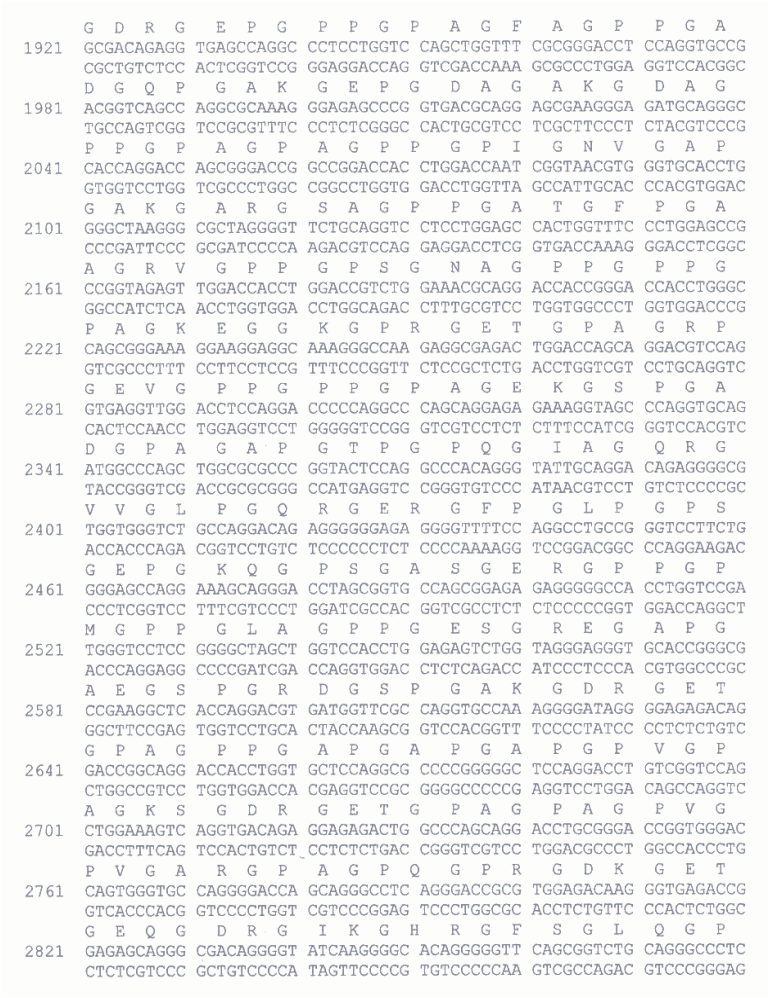

Figure 2.3: Codon optimized collagen sequence, part 3

Figure 2.4: Codon optimized collagen sequence, part 4

Table 2.1: Intermediate vectors generated for the assembly of the collagen 1 gene. Fragment correspond to the fragment amplified with homonymous primers

Fragment	Length (base pairs)	Vector	Oligos added to the mixture
5'sp- 3'2	750	pUC18(5'sp- 3'2)	col1-col8/ 59R-67R
5'2- 3'5	1400	pUC18(5'2- 3'5)	col4-col19/ 49R-62R
5'5- 3'8	1356	pUC18(5'5- 3'8)	col15-col30/ 38R-52R
5'8- 3'10	876	pUC18(5'8- 3'10)	col25-col33/ 34R-42R
5'sp- 3'5	2150	pUC18(5'sp- 3'5)	N/A
5'5- 3'10	2232	pUC18(5'5- 3'10)	N/A

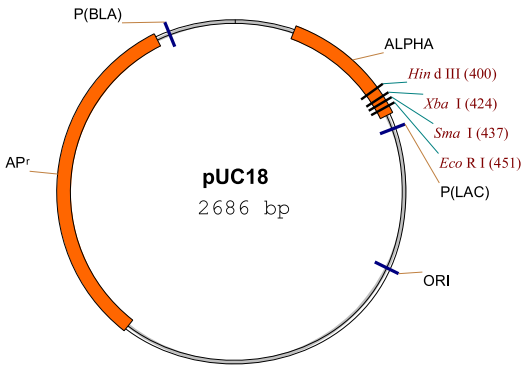

Figure 2.5: Plasmid pUC18 used as intermediate vector for assembly of the gene

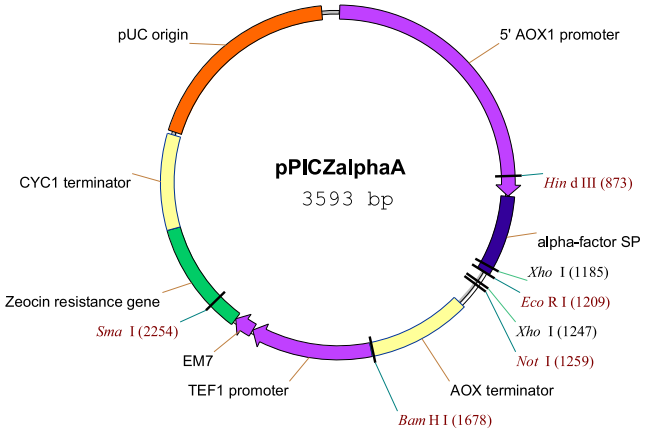

Figure 2.6: Plasmid pICZαA used as a vector for expression of proteins in *Pichia pastoris*. Selection of positive transformants was based on resistance to Zeocin®

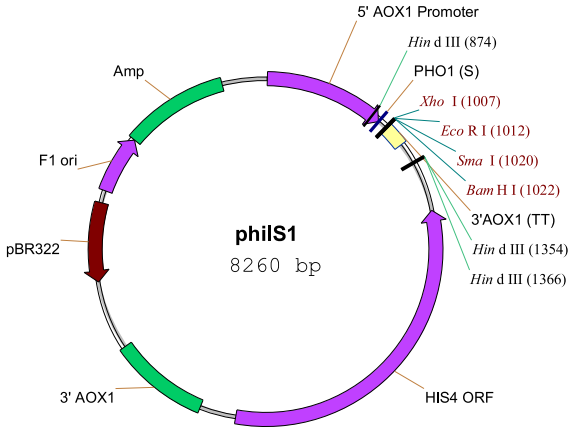

Figure 2.7: Plasmid pHIL-S1 used for transformation of *Pichia pastoris* strain GS115. Selection of positive transformants was based on the ability to grow on a histidine-deficient medium

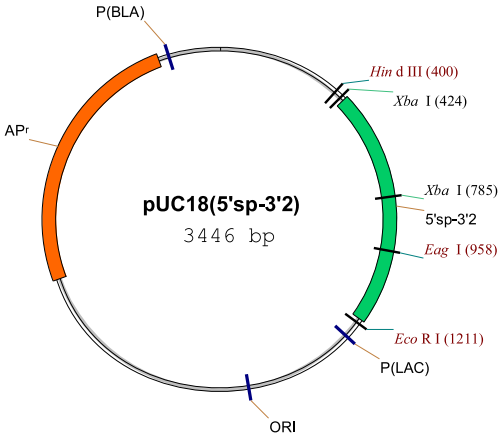

Figure 2.8: Intermediate vector pUC18(5'sp-3'2)

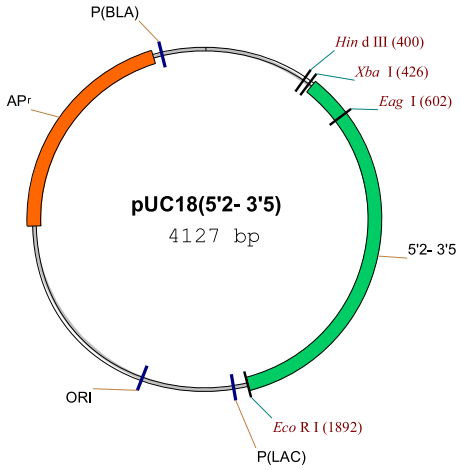

Figure 2.9: Intermediate vector pUC18(5'2-3'5)

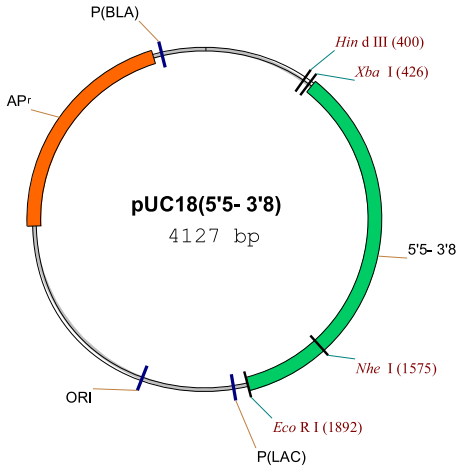

Figure 2.10: Intermediate vector pUC18(5'5-3'8)

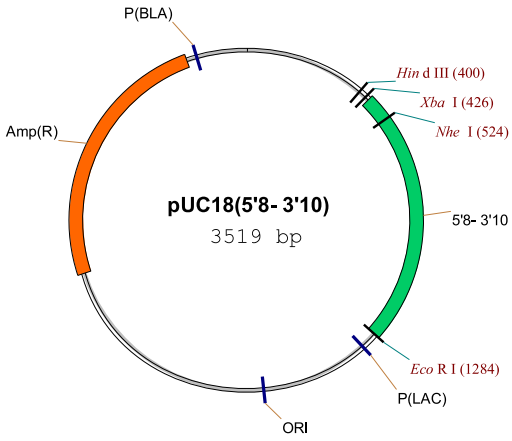

Figure 2.11: Intermediate vector pUC18(5'8-3'10)

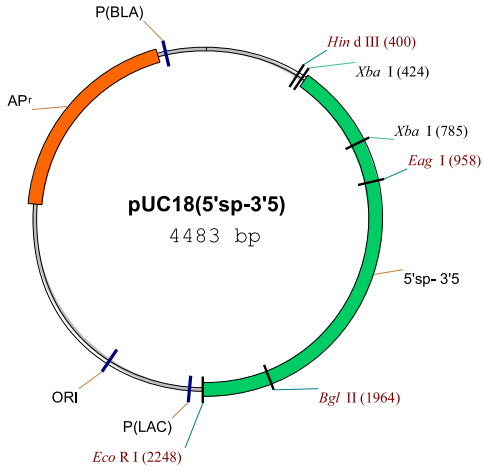

Figure 2.12: Intermediate vector pUC18(5'sp-3'5) generated by ligation of 5'sp-3'2 and 5'2-3'5 using restriction enzyme *Eag*I

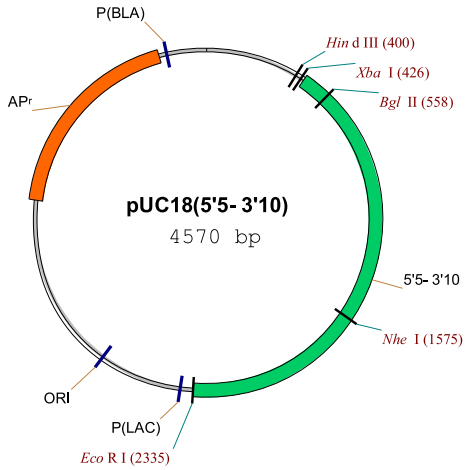

Figure 2.13: Intermediate vector pUC18(5'5-3'10) generated by ligation of corresponding fragments of 5'5-3'8 and 5'8-3'10 using restriction enzyme *Nhe*I

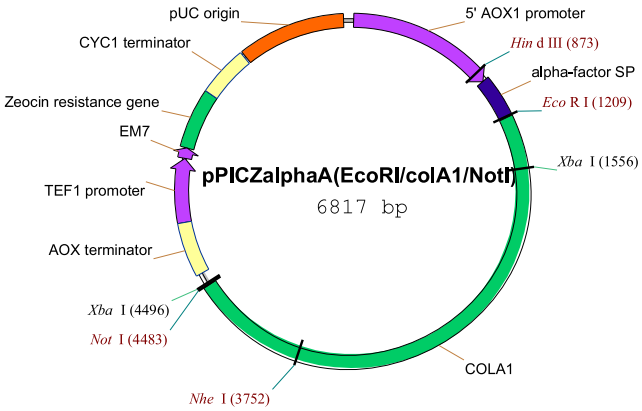

Figure 2.14: Expression vector pICZαAEcoRI-COLA1-NotI

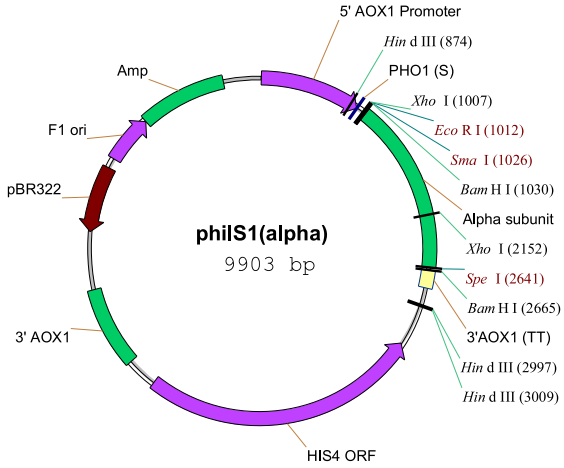

Figure 2.15: Expression vector pHIL-S1(α)

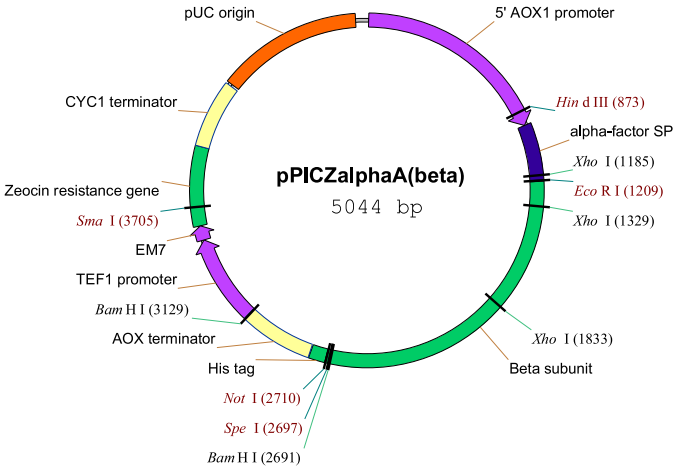

Figure 2.16: Expression vector pICZαA(β)

Table 2.2: Plasmids and strains generated for the expression of COL1A1 and prolyl 4-hydroxylase.

Expression vector	Strain	Selection	Polypeptides expressed
pPICZα A(proCOLA1)	proCOLA1	Zeocin®	proalpha 1 chain of type I procollagen
pPICZalpha A(P4H-beta)	beta	Zeocin®	P4H- beta subunit
pHIL- S1(alpha)	alpha	His+	P4H- alpha subunit
pPICZalpha A(P4H-beta), pHIL- S1(alpha)	alpha + beta	Zeocin®, His+	P4H- alpha subunit, P4H- beta subunit

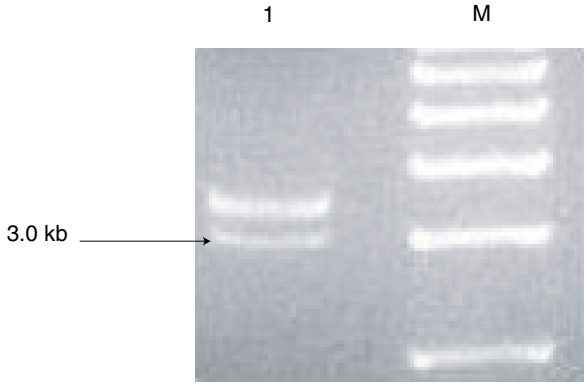

Figure 2.17: Agarose gel stained with ethidium bromide to visualize the gene encoding for (α1) chain of collagen type I. The gene was assembled starting from oligonucleotides that were ligated, amplified by PCR and cloned into intermediate vectors to yield the 3262 base pairs corresponding to the gene. Lane 1, pICZA(COL1A1) digested with *Xba*I and *Not*I showing the gene (3kb) and the vector fragment; lane M, DNA marker.

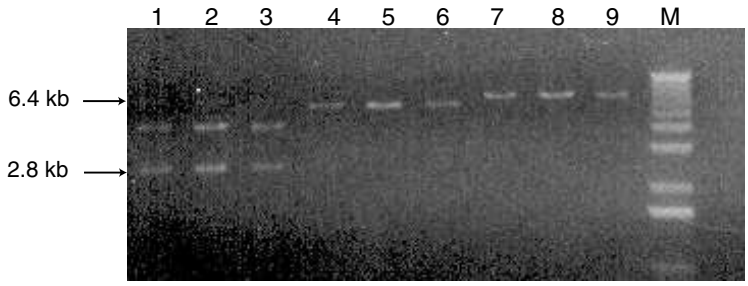

Figure 2.18: Agarose gel stained with ethidium bromide to visualize plasmid DNA isolated from DH5α positive colonies transformed with expression vector pICZαA*Eco*RI-COL1A1-*Not*I. Plasmid DNA was isolated from positive colonies growing on Zeocin® containing medium and subjected to restriction enzyme analysis. Lanes 1, 2 and 3: colonies digested with *Nhe*I and *Eco*RI; lanes 4, 5, and 6: colonies digested with *Nhe*I and *Not*I; lanes 7, 8 and 9 digested with *Eco*RI; lane M, DNA marker.

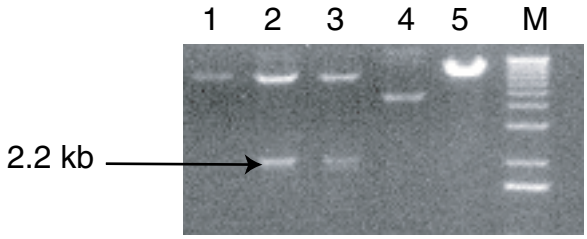

Figure 2.19: Agarose gel stained with ethidium bromide to visualize plasmid DNA isolated from DH5α positive colonies transformed with expression vector pHIL-S1(α). Plasmid DNA was isolated from positive colonies growing on ampicillin medium and subjected to restriction enzyme analysis.Lanes 2 and 3 DNA digested with *Hin*dIII; lane 4 negative plasmid control; lane 5 positive plasmid control; lane M DNA marker

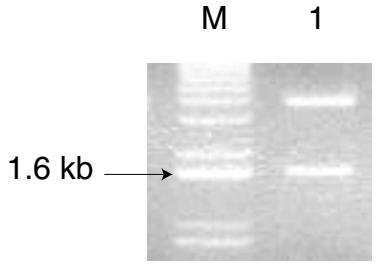

Figure 2.20: Agarose gel stained with ethidium bromide to visualize plasmid DNA isolated from DH5α positive colony transformed with expression vector pICZαAβ. Plasmid DNA was isolated from a positive colony that grew on Zeocin ® medium and subjected to restriction enzyme analysis. Lane 1, colony digested with *Eco*RI and *Not*I; lane M, DNA marker.

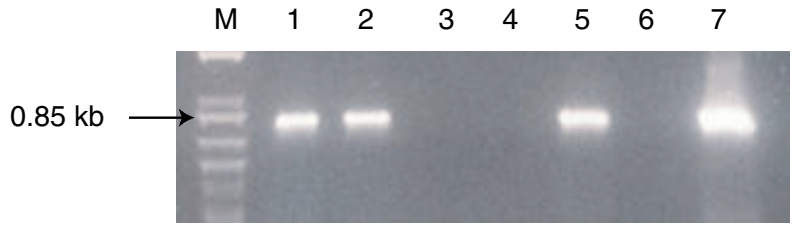

Figure 2.21: Agarose gel stained with ethidium bromide to visualize DNA fragment amplified from *Pichia pastoris* GS115 transformed with pICZαA*Eco*RI-COL1A1-*Not*I. Expression plasmid DNA was isolated, concentrated and linearized with *Sal*I. GS115 strain was transformed by the Easy Comp method (Invitrogen, Carlsbad, CA). Colonies were selected on YPD plates containing Zeocin® and screened by direct-PCR. Lane M, marker; lanes 1, 2, 3, 4 and 5 corresponds to the numbers assigned to the colonies tested; lane 6 empty plasmid used as negative control; lane 7, plasmid DNA used as positive control

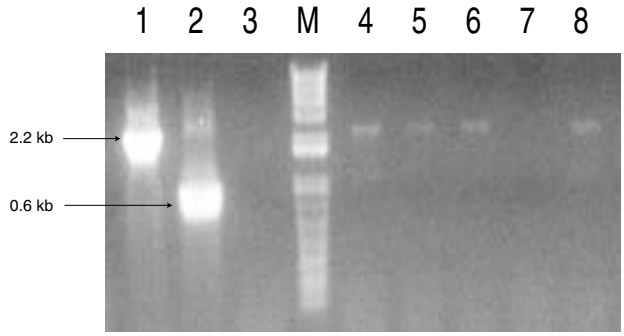

Figure 2.22: Agarose gel stained with ethidium bromide to visualize DNA fragment amplified from *Pichia pastoris* GS115 transformed with from *Pichia pastoris* GS115 transformed with pICZαAβ. Expression plasmid DNA was isolated and the parental strain was transformed as described in Methods. Positive colonies were selected on YPD plates containing Zeocin$^{\textcircled{R}}$ and screened by direct PCR. Lane 1, plasmid DNA used as positive control; lane 2, empty plasmid used as negative control; lane 3, water control; lane M, DNA marker; lanes 4, 5, 6, and 8 positive colonies.

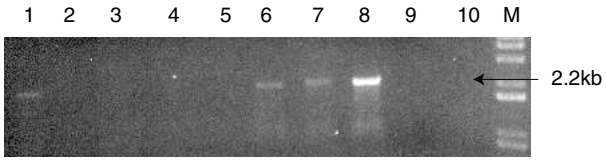

Figure 2.23: PCR screening of *Pichia pastoris* GS115 transformed with pHIL-S1(α). Expression plasmid DNA was isolated and the parental strain was transformed as described in Methods. Positive colonies were selected on histidine deficient plates. Lanes 1, 6, 7 positive yeast colonies, lane 8 positive control; lane M, marker

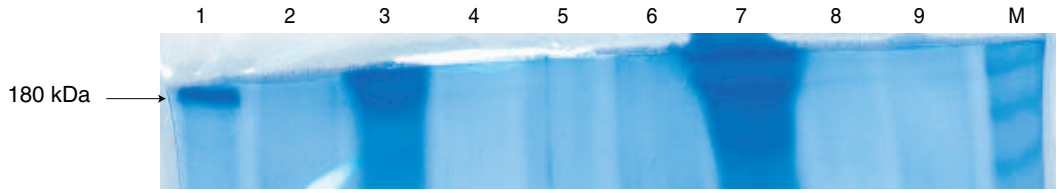

Figure 2.24: Coomassie-stained SDS-PAGE showing expression of homotrimeric procollagen I in *Pichia pastoris*. 1.5 ml of culture medium was concentrated to 70μl; 40μl were boiled for 3 minutes with 6x loading buffer and then 20 μl were loaded into each well of a 10% SDS gel. 100volt were applied for 2 hours and the gel was then stained overnight with Coomassie blue. Lane 1, *Pichia*-derived non-hydroxylated collagen I; lane 2, negative control; lane 3, concentrated soluble fraction from cell lysate; lane 7, concentrated culture medium; lane M, marker

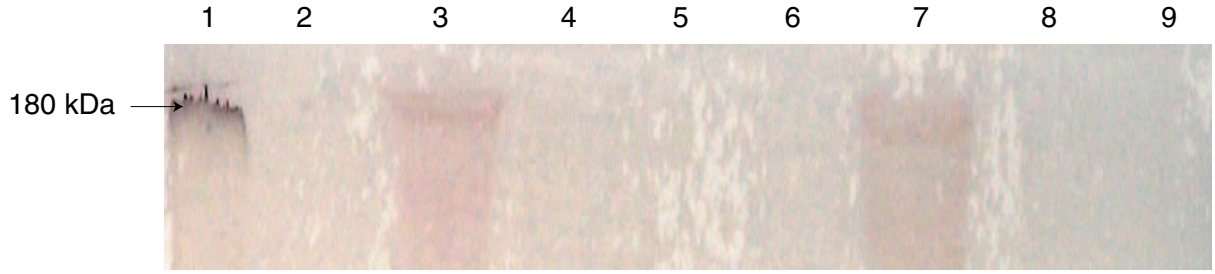

Figure 2.25: PVDF membrane stained with specific antibodies to show expression of homotrimeric procollagen I in *Pichia pastoris* Western blotting of concentrated growth medium containing recombinant collagen I as described above. The samples were transferred with 18volt for 40 min to a nitrocellulose membrane that was incubated in the blotting solution for 2 h, and then incubated overnight with specific antibodies that recognize a collagen domain. After washing, the membrane was incubated with secondary antibody for 3 h. Lane 1, *Pichia*-derived non-hydroxylated collagen I; lane 2, negative control; lane 3, concentrated soluble fraction from cell lysate; lane 7, concentrated culture medium; lane M, marker.

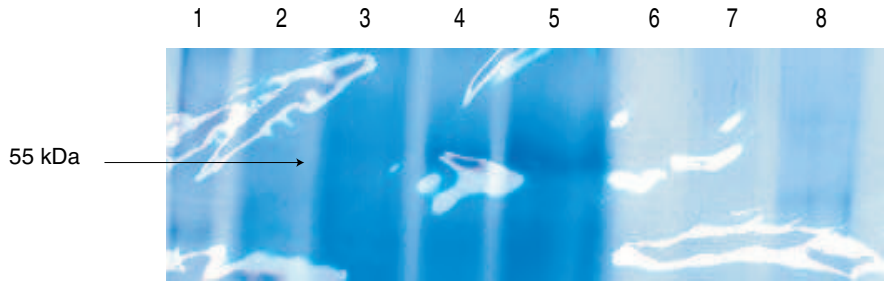

Figure 2.26: Coomassie-stained SDS-PAGE showing expression of β subunit of prolyl 4-hydroxylase in *Pichia pastoris*. 1.5ml of culture medium was concentrated to 70μl, 40μl were boiled for 3 min with 6x loading buffer and then 20μl were loaded into each well of a 10% SDS gel. One hundred volt were applied for 2 h and the gel was then stained overnight with Coomassie blue. Lane 1, marker; lanes 4 and 5 concentrated culture medium.

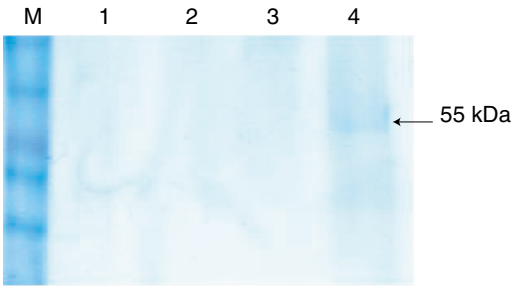

Figure 2.27: Coomassie-stained SDS-PAGE showing expression of α and β subunit of prolyl 4-hydroxylase in *Pichia pastoris*. 1.5ml of culture medium was concentrated to 70μl, 20μl were boiled for 3 min with 6x loading buffer and then loaded into each well of a 10% SDS gel. One hundred volt were applied for 2 h and the gel was then stained overnight with Coomassie blue. Lane M, molecular marker; lane 1, negative control; lanes 2, 3, and 4 concentrated culture medium.

Chapter 3

Collagen expression in barley

3.1 Abstract

Several recombinant expression systems have been developed in order to produce collagen. Among them, tobacco has been proven to correctly produce hydroxylated collagen I, opening the possibilities for expression in barley that has been shown to be suitable for the production of recombinant proteins. Towards this goal, vectors were developed carrying the genes needed for the synthesis of homotrimeric hydroxylated procollagen I in the grain of barley. In order to test the hypothesis that barley can correctly express and fold procollagen, barley was transformed with plasmids carrying the genes for (α)1 chain of collagen type I and both α and β subunits of prolyl 4-hydroxylase. Each gene was under the control of the barley hordein D-promoter and signal peptide to target the peptide to the storage protein vacuoles of the endosperm in maturing barley grains. Resistance to bialaphos driven by the bar gene was used as a selectable marker, and immature zygotic embryos were transformed with *Agrobacterium tumefa-*

ciens carrying the desired plasmid. Green plants are being generated and will be tested by polymerase chain reaction with specific primers coding for an internal fragment of the collagen gene and the prolyl 4-hydroxylase gene.

3.2 Introduction

Several expression systems have been developed to express heterologous proteins. Among them, mammalian and plant systems seem to be the most suitable to produce recombinant human proteins with the correct structure.

Synthesis of recombinant collagen presents several challenges because of the number of post-translational modifications, requiring eight different enzymes, some of them unique to collagens; all of them are needed to achieve a fully folded, triple helical conformation. Procollagen produced in insect cells and yeast must be treated with pepsin [*Bulleid et al.,* 2000], a process that could damage the telopeptides [*Bulleid et al.,* 2000] and, as a consequence, pepsin-extracted collagen might not form fibrils [*Leibovich and Weiss,* 1970*; Bulleid et al.,* 2000]. However, these systems provide the basic knowledge needed to develop new expression systems that can successfully process collagen.

Expression in mammalian cell lines

Mammalian cell lines have been engineered to produce recombinant collagen, resulting in procollagen secreted into the culture medium [*Fichard et al.,* 1997]. In this experi-

ment, human embryonic kidney cells (293-EBNA) were transfected with the full-length human α1 chain of collagen V using an episomal vector. High yields (15 mg/ml) of recombinant collagen were secreted into the culture medium. In the presence of ascorbate, the α1(V) collagen was correctly folded into a stable triple helix as demonstrated by electron microscopy and pepsin treatment. Circular dichroism data confirmed the triple-helix conformation and indicated a melting temperature of 37.5°Cfor the recombinant homotrimer. The major secreted form was a 250-kDa polypeptide (a1FL).

Expression in mice

The mammary gland is a promising expression system, and several proteins have already been expressed satisfactorily from different species including mice [*Prunkard et al.,* 1996], rabbit [*Stromqvist et al.,* 1997] and pig [*Paleyanda et al.,* 1997]. The system uses mammary-gland-specific promoters that drive the expression of foreign proteins in milk. In the case of collagen, two reports have been published describing the use of transgenic mice for collagen production. The first one [*John et al.,* 1999] involved secretion of a truncated ((α)1)3 molecule using the αS1-casein promoter. The transgenic mouse lines were generated to also express prolyl 4-hydroxylase, but the melting temperature was lower than bovine-extracted collagen, possibly because the chain had a lower percentage of hydroxyproline residues than the full-length chain.

In the second report [*Toman et al.,* 1999], transgenic mice were generated containing the α S1-casein mammary-gland-specific promoter operatively linked to 37kb of

the human α1(I) procollagen structural gene and 3' flanking region. The frequency of transgenic lines established was 12%. High levels of soluble triple helical homotrimeric [(α1)3] type I procollagen were detected (up to 8 mg/ml) exclusively in the milk of six out of nine lines of lactating transgenic mice. The transgene-derived human procollagen chains underwent efficient assembly into a triple helical structure. Although proline or lysine hydroxylation is not common for milk proteins, procollagen was detected with these post-translational modifications. The procollagen was stable in milk, and minimal degradation was observed.

Both results showed that the mammary gland is capable of expressing a large procollagen gene construct, efficiently assemble the individual polypeptide chains into a stable triple helix, and secrete the intact molecule into the milk.

Expression in silkworms

Transgenic silkworms are able to produce collagenous domains and secrete them into the cocoons [*Tomita et al.,* 2003]. A mini-chain encoding for type III collagen was used as transgene, under the control of a fibroin light chain promoter and an enhanced green fluorescent protein (EGFP) coding region as a selectable marker. The cDNA was inserted between the fibroin L-chain gene 5'-flanking and 3'-flanking sequences. The expression units were inserted into a vector containing the gene for red fluorescent protein. Pre-blastoderm embryos were injected with the constructs and allowed to develop at 25°C. The construct was expressed, but due to lack of hydroxylation, the

protein was not assembled in the silk gland.

Expression in tobacco

Ruggeiro and collaborators, [*Ruggiero et al.,* 2000], using tobacco as an expression system, obtained primary transformants from different constructs encoding human proα1(I) chain. As expected, the plants produced homotrimer molecules, but due to the specificity of the plant prolyl 4-hydroxylase, the plant enzyme cannot ensure the prolyl-hydroxylation of the collagenous tripeptide X-Pro-Gly [*Tanaka et al.,* 1981]. On the other hand, since plants contain PDI, that corresponds to the beta subunit of P4H, it might act as a chaperon during chain assembly [*Ruggiero et al.,* 2000], aiding in the formation of homotrimers. In a different report [*Perret et al.,* 2001], demonstrated that the amount of proline-hydroxyproline is a limiting factor in unhydroxylated procollagen assembly.

In a later study, Merle and collaborators ([*Merle et al.,* 2002]), produced hydroxylated homotrimer collagen I as a result of *Agrobacterium tumefaciens* transformation of intact tobacco leaves. Plants expressing collagen, PDI and α subunit were analyzed and the results shown conclusively that the system produced thermally stable triple helical homotrimeric collagen I. The level of hydroxyproline formation was identical to the level obtained with *Saccharomyces cerevisiae*, but in a lower concentration compared to that achieved with *Pichia pastoris*.

Barley can produce and store recombinant proteins in the endosperm up to 1g per

kilo of grain [*Horvath et al.,* 2000]. Several proteins have been successfully expressed in the barley endosperm including lysozyme, lactoferrin, human serum albumin and an engineered thermostable beta-glucanase using *Agrobacterium tumefaciens*-mediated transformation of immature zygotic embryos [*Horvath et al.,* 2000*; Stahl et al.,* 2002].

For my study, transcription of the collagen gene was carried out with the hordein D-promoter and the mRNA translated by the polysomes of the endoplasmic reticulum of the endosperm of the barley grain. Endosperm specific expression of proteins driven by the hordein promoter are stably inherited in the T1, T2 and following generations, and recombinant proteins segregate in a Mendelian fashion [*Cho et al.,* 2002*; Horvath et al.,* 2000].

The genes coding for collagen and both subunits of prolyl 4-hydroxylase were fused to the nucleotide sequence of the hordein 3 (hor-3) signal peptide. Hor3 signal peptide contains 21 amino acids that direct the nascent polypeptide chains to the lumen of the endoplasmic reticulum where the signal peptide is cleaved off during or immediately after translocation. From there the proteins are delivered via the Golgi apparatus into the storage vacuoles of the endosperm [*Horvath et al.,* 2000]. Deposition of the proteins in the storage protein bodies protect them from programmed cell death during the maturation phase of the grain.

The single-cassette vector also has the *bar* gene of *Streptomyces hygroscopicus* under the control of the maize ubiquitin promoter and nos terminator for selection of positive transformants on bialaphos-containing medium. The *bar* gene codes for

the enzyme phosphinothricin acetyl transferase (PAT), which inactivates the toxicity of L-phosphinothricin (L-PPT). This later compound inhibits the enzyme glutamine synthetase, an inhibition that results in a high accumulation of ammonium. This ammonium interferes with the electron transport in chloroplasts and mitochondria, causing cell death. In transformed plants, PAT detoxifies L-PPT by acetylation and, as a result, calli carrying the *bar* gene will survive and regenerate into green plants [*De Block et al.,* 1995].

The main goal of the work reported in this chapter is to provide the basis for the production of transgenic barley plants expressing the genes needed for hydroxylated collagen. In order to achieve this goal, it was necessary to develop the vectors and transform immature zygotic embryos of barley.

Specifically, the following objectives were addressed:

- Develop the vectors needed for the *Agrobacterium tumefaciens* transformation of barley
- Transformation of immature embryos
- Production of barley plants containing the genes for the production of hydroxylated homotrimeric procollagen I.

3.3 Materials and Methods

Plasmids

Four different plasmids were used for the development of expression vectors suitable for *Agrobacterium tumefaciens*-mediated barley transformation. The characteristics of each one are detailed as follows.

Plasmid pUC18 from MBI Fermentas (Hannover, MD) was used as an intermediate cloning vector. It is a high copy-number plasmid that has the pMB1 replicon responsible for the replication of the plasmid source and the *bla* gene that confers resistance to ampicillin (Fig. 2.11).

Plasmid pHordSpNos was derived from pUC18. It has the gene that codes for ampicillin resistance. On the multiple cloning site it has the hordein D-promoter plus signal peptide code and the nos terminator for the generation of intermediate vectors (see Fig. 3.1).

Plasmid RS366 was also derived from pUC18. It contains the hordein-D promoter flanked by *Hind*III and *Nco*I restriction sites for easy cloning procedures (see Fig. 3.2).

Plasmid pJH260 is a single cassette cloning vector derived from the binary vector pJH20 (Jintai Huang, unpublished) by a three way ligation of a *Hind*III-*Sma*I fragment containing the bar gene 3' to the ubiquitin promoter and a *Sma*I-*Eco*RI fragment containing the Nos terminator from pUBARN into *Hind*III-*Eco*RI digested pJH20. It has the gene for kanamycin resistance (Fig. 3.3).

Strains

Escherichia coli strain DH5α: cloning of the desired plasmid was performed into *E. coli* strain DH5 (α sup E44Δlac U169 (Φ80 lacZΔM15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1).

Agrobacterium tumefaciens AGL-1 carries the disarmed Ti plasmid that has the hypervirulence genes attenuated and also harbors the *rif* gene for resistance to rifampicin.

Barley vector construction

A fragment encoding 250 base pairs of the hordein D-promoter was amplified from plasmid RS366 using primers 5'NcoI-prom and 3'col-sp that have 8 base pairs that overlap the collagen gene. Fragment (5'sp-3'2)[Fig. 2.8] was amplified using 5'sp-col and 3'2 [Table 2.1] from the original ligation mixture using *Pfu* polymerase and the PCR program previously described. The two fragments were spliced together following the SOE-PCR procedure described by Horvath and collaborators [*Horvath et al.,* 2000]. The resulting fragment was cloned into pUC18 and sequenced (Amplicon Express, Pullman, WA). One error was found and it was corrected by replacing part of the fragment using *Xba*I enzyme. The fragment encoding (NcoI-signal peptide-procol1A1-XbaI) was inserted into pICZA(COL1A1) [p.27, cf Fig. 2.17] and digested with the same enzymes; the resulting fragment was *Not*I-blunted, *Nco*I-digested, and ligated into RS366 previously digested with *Eco*RI, blunted, purified and digested with

*Nco*I. Five microliters of the ligation mixture were used to transform DH5α-competent cells, and positive colonies were plated out on LB ampicillin plates. Transformants were confirmed by PCR and restriction digests.

To add the NOS terminator at the end, the plasmids pHordSpNos and RS366(COL1A1) were digested with *Hind*III and *Sac*I, and then ligated to generate plasmid pHorSp(COL1A1)Nos (Fig. 3.6).

Plasmid pJH260 is a single cassette vector that contains the bar gene used for selection against bialaphos

The fragment encoding hordein-D promoter-signal peptide-COL1A1-nos terminator sequence was digested from plasmid pHorSp(COL1A1)Nos with *Eco*RI, blunted and digested with *Hind*III, and cloned into pJH260 previously *Spe*I-blunted and *Hind*III-digested to yield pCO200 (see Fig. 3.8). DH5α cells were transformed with the ligation, and then kanamycin-resistant colonies were selected and analyzed by PCR and restriction enzyme digestion. Minipreps were prepared with the BioRad miniprep kit.

A similar procedure was followed to generate the intermediate vectors containing the α and β subunits for prolyl 4-hydroxylase. For the α subunit, a fragment encoding the hordein D-promoter was amplified from plasmid RS366 using primers 5'NcoI-prom and 3'sp-alpha. Using program COL2(25) (Appendix, Table 5.2), 8 overlapping base pairs were added to the α subunit amplified from plasmid 002 (Fibrogen, San Francisco, CA) using primers 5'sp-alpha and 3'alpha. Both fragments were spliced together with SOE-PCR as previously described. The fragment was cloned into pUC18 previously

digested with *Sma*I and dephosphorylated, and then sequenced. No errors were found and the plasmid pUC18(NcoI-alpha) was digested with *Spe*I, blunted and digested again with *Nco*I. The resulting fragment was cloned into RS366 digested with *Eco*RI, blunted and digested with *Nco*I. This intermediate vector (RS366alpha) was digested with *Hin*dIII and *Eco*RI and cloned into pHorspnos digested with the same enzymes to yield plasmid pHorspαnos (Fig. 3.4).

A similar protocol was followed for the generation of the intermediate plasmid pHorspβnos. The amplification of the fragment corresponding to part of the hordein promoter was done using primers 5'NcoI-prom and 3'sp-beta taking as a template plasmid RS366. The gene coding for the β subunit was amplified from plasmid 003 (Fibrogen, San Francisco, CA) using primers 5'sp-beta and 3'beta. Both fragments were spliced together, cloned into pUC18 and sequenced as described. The resulting plasmid [pUC18(NcoI-beta)] was digested as previously described for pUC18(NcoI-alpha) to yield plasmids RS366beta and pHorspβnos (Fig. 3.5).

pHorspαnos and pHorspβnos were treated with *Hind*III/*Eco*RI adaptor (Easy-clones systems) and ligated into pCO200 to yield pCO210 and pCO220, respectively, with protocol as described by the manufacturer (Figs. 3.9 and 3.10). Colonies were selected for kanamycin resistance.

To generate a plasmid containing COL1A1 and α and β subunits of prolyl 4-hydroxylases, pCO200 was digested with *Hind*III and dephosphorylated. Concurrently, plasmids pHorspαnos and pHorspβnos were *Hind*III and *Eco*RI digested, and

a three way ligation was done to produce plasmid pCO250 (Fig. 3.11).

Agrobacterium tumefaciens strain AGL-1 was electroporated at 1.25kv, 25F and 200Ω on the Bio-Rad Gene Pulser; kanamycin resistant colonies were then selected, and transformation was confirmed by restriction site analysis.

Barley transformation was done as described in Horvath et al [*Horvath et al.,* 2000], with small modifications. Time-line of transformation is shown in Table 3.1. Embryo axes were removed by cutting. Three different experiments were conducted (Table 3.2). Co-cultivation with the corresponding *Agrobacterium* strain was done for 40 minutes, and the embryos were then transferred to CIM(0) plates and maintained at 24°Cfor 48 hours before transferring into CIM(4) media.

DNA extraction and PCR screening of positive transformants

At the moment that the putative transformants are transferred to soil from root generating media, a piece of leaf was cut from each plant and put into an Eppendorf tube. The tissue was frozen by immersion into liquid nitrogen and stored at -20°C. DNA from each putative transformant was then isolated as follows. To each Eppendorf tube containing leaf tissue, 400μl of extraction buffer composed of 200mM Tris-HCl ,pH 8, 250mM NaCl, 25mM EDTA and 0.5%SDS was added and the tissue was homogenized with a grinder. To the homogenized tissue, 400μl of chloroform was added, vortexed and spun at maximum speed for 10 minutes. Three hundred microliters of the supernatant were transferred to a clean tube, and the DNA was precipitated by addition

of 300μl of isopropanol and then incubating the tubes at room temperature for 1 h. Samples were centrifuged for 5 minutes at maximum speed and the resulting pellet was washed twice with 75%EtOH, dried and resuspended in 50μl of sterile water.

Reactions were carried out to screen for the presence of the gene (Table 3.2). The diagnostic samples were amplified from the template with a 25μl mixture that contained 1μl of plant DNA, 20 pmol of each primer, 2mM dNTPs, *Pfu* buffer and *Pfu* polymerase as suggested by the manufacturer. PCR parameters were the same as described in table 5.2.

3.4 Results and Discussion

3.4.1 Vector construction

Four single cassette vectors were made available for barley transformation. Each gene was under the control of the hordein D-promoter that effects transcription of the gene in the endosperm (see Fig. 3.7). Translation of the protein precursor with its signal peptide leads to transfer into the lumen of the endoplasmic reticulum and from there the mature protein is transferred into the storage vacuoles [*Cameron-Mills et al.*, 1978; *Cameron-Mills and von Wettstein*, 1980]. Plasmid pCO200 has the gene coding for the α1 chain of collagen I flanked by the hordein D-promoter plus signal peptide sequence and nos terminator. Plasmids pCO210 and pCO220 were generated from plasmid pCO200 and also have the gene coding for α and β subunits of pro-

lyl 4-hydroxylase respectively. The restriction analysis of plasmids pCO200, pCO210, pCO220 can be seen in figure 3.12 after digestion with *Not*I and *Hin*dIII. The 4.2kb fragment correspond to COL1A1 flanked by hordein promoter and nos terminator; the 3.0kb fragment corresponds to the plasmid fragment from pJH260. The 2.4kb and 2.3kb fragments in lanes 3 and 4 correspond to α and β subunits of prolyl 4-hydroxylase flanked also by hordein promoter and nos terminator, respectively. For the production of hydroxylated homotrimeric collagen I, a plasmid was constructed containing the three genes (α1 chain, α and β subunits of prolyl 4-hydroxylase), each flanked by its own hordein D-promoter plus signal peptide sequence and nos terminator. Restriction analysis conducted to the plasmid are conclusive about integration and correct orientation of the fragments corresponding to the α and β subunits of P4H. The analysis can be seen in Fig. 3.13. Lane 1 shows the plasmid digested with *Hin*dIII, that corresponds to α and β subunits of prolyl 4-hydroxylase. Lane 2 corresponds to the plasmid digested with *Eco*RI that shows evidence of the correct cut of the plasmid. In lane 3 it can be seen the plasmid digested with *Hin*dIII and *Eco*RI. The fragments corresponding to COL1A1 and the α and β subunits of prolyl 4-hydroxylase are recognizable as bands with a size of 4.1, 2.4 and 2.3kb. The bands with a size of 1.3 and 1.4kb corresponds to fragments of the plasmid pJH260. This plasmid was also used for transformation of immature zygotic embryos.

3.4.2 Barley transformation

In one experiment, 95 embryos were co-cultivated with plasmid pCO200. In subsequent transformation experiments, 600 embryos were co-cultivated with a mixture of *Agrobacterium tumefaciens* containing plasmid pCO210 or pCO220. It was expected that stable transgenic plants would be generated that expressed the three genes from pCO210 and pCO220 in single primary transformed scutellum cells. Regeneration of transgenic plants from such cells would then provide transformants producing the protein from all three genes. Such simultaneous expression from two plasmids has been demonstrated by Merle and collaborators [*Merle et al.,* 2002]. Transformations are also being carried out with plasmid pCO250; these will lead to transgenic plants expressing the three genes needed for the production of hydroxylated homotrimeric collagen type I. Embryos were selected on callus-induction medium containing bialaphos for selection of positive transformants [*De Block et al.,* 1995]. Timentin was used to inhibit the growth of *Agrobacterium* on the plate. Timentin is composed of a mixture (15:1) of ticarcillin and clavulanic acid. Since *Agrobacterium tumefaciens* has β-lactamase activity, clavulanic acid acts as a competitive inhibitor of β-lactamase, resulting in the death of bacteria due to cell wall lysis by ticarcillin action [*Nauerby et al.,* 1997].

At this point, three plants transformed with pCO200 and three plants transformed with a mixture of pCO210 and pCO220 were tested with specific primers with negative results. This is not surprising because of the level of transformation for barley cultivar Golden Promise is often not more than 4% [*Tingay et al.,* 1997].

It is expected that among the additional plants that are being generated, positive transformants will be found.

3.5 Tables and Figures

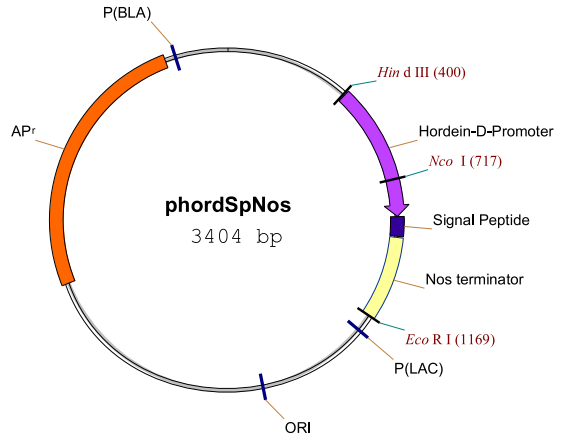

Figure 3.1: Intermediate vector pHorSpNos

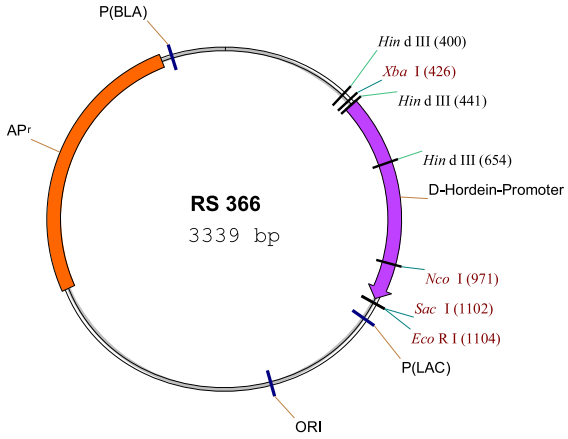

Figure 3.2: Intermediate vector RS366

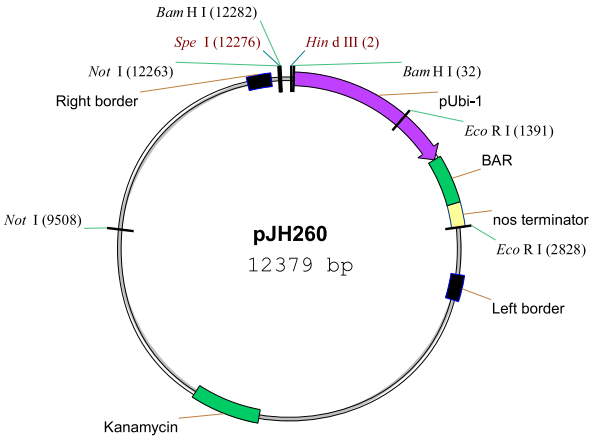

Figure 3.3: Plasmid pJH260 with the herbicide resistance gene (bar) under the control of ubiquitin promoter and nos terminator used as selectable marker for putative transformants.

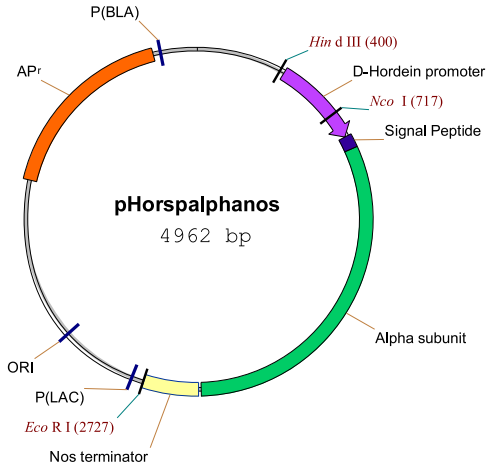

Figure 3.4: Intermediate plasmid pHorspαnos

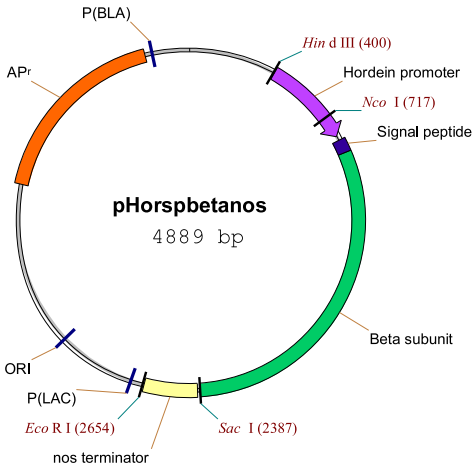

Figure 3.5: Intermediate plasmid pHorspβnos

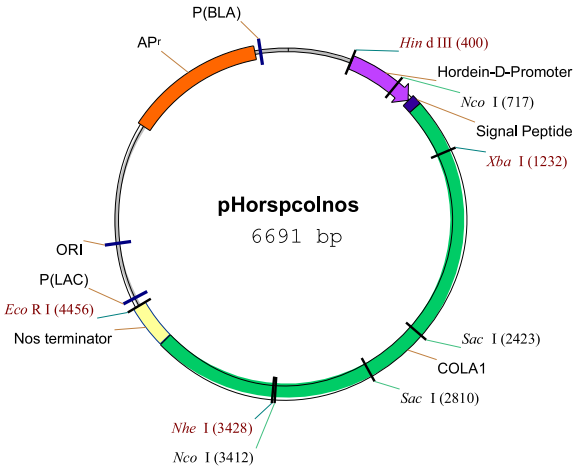

Figure 3.6: Intermediate plasmid pHorspCOL1A1nos

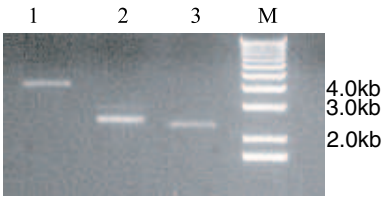

Figure 3.7: Agarose gel stained with ethidium bromide showing plasmid DNA obtained after addition of hordein D-promoter plus signal peptide and nos terminator to the genes coding for COL1A1, α and β subunits of P4H. Hordein promoter plus signal peptide was added at the beginning of each gene and nos terminator at the end. Lane 1, hordein D-promoter plus signal peptide-COL1A1-nos terminator; lane 2, hordein D-promoter plus signal peptide-α-nos terminator; lane 3, hordein D-promoter plus signal peptide-β-nos terminator; lane M, DNA marker.

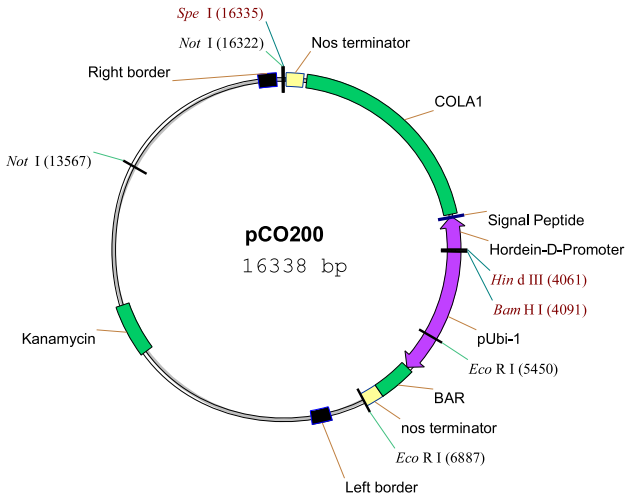

Figure 3.8: Plasmid pCO200 encoding gene for α1 chain of collagen type 1 under the control of hordein 3-D promoter, signal peptide and Nos terminator.

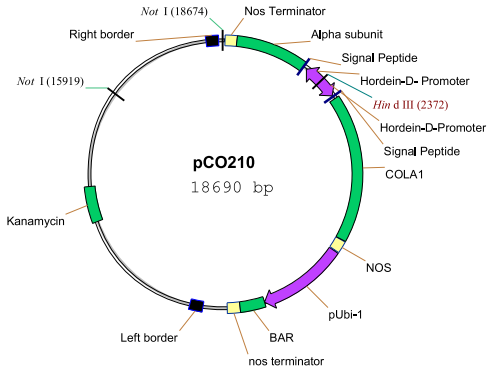

Figure 3.9: Plasmid pCO210 encoding collagen 1 gene and α subunit of P4H. Each gene is under the control of its own hordein 3-D promoter, signal peptide and Nos terminator.

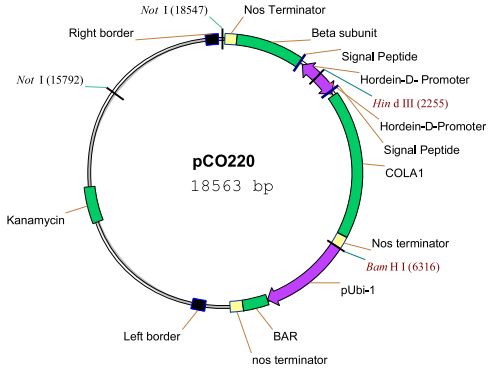

Figure 3.10: Plasmid pCO220 encoding collagen 1 gene and β subunit of P4H. Each gene is under the control of its own hordein 3-D promoter, signal peptide and Nos terminator.

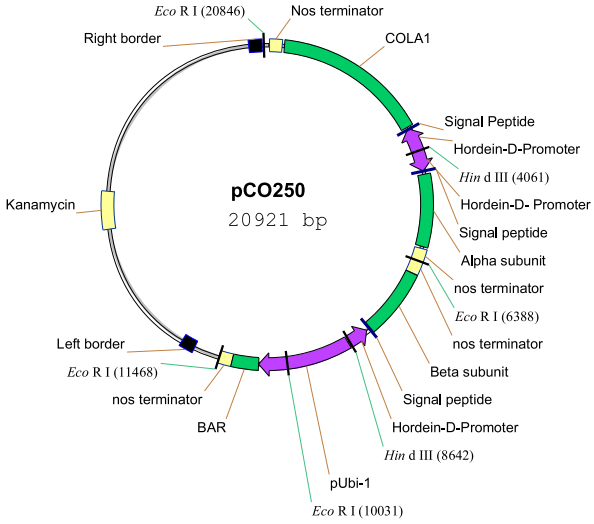

Figure 3.11: Plasmid pCO250 encoding for collagen 1 gene, α and β subunit of P4H. Each gene is under the control of its own Hordein 3-D promoter, signal peptide and Nos terminator

Table 3.1: *Agrobacterium tumefaciens* time-line for transformation.

Day/ Week	Procedure
1	Start *Agrobacterium* liquid culture and grow at room temperature with shaking at 100 rpm
2	Disinfect and cut immature barley embryos. Place on CIM(0) plates and inoculate with liquid culture for 40 minutes. Transfer the embryos to a new CIM(0) plate and keep them in dark at 24°C
4	Transfer the embryos to a CIM(4) plate and cultivate in dark at 24°C
week 3	Transfer to second round of CIM(4) selection
week5	Transfer to SGM and cultivate plantlets at 16 hour light/ 8 hours dark at 24°C
week 11 to 16	Transfer to RGM
Week 16 to 20	Transfer to soil and take DNA samples
Month 7	Harvest mature seeds

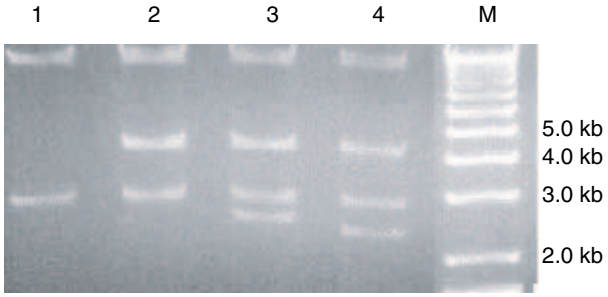

Figure 3.12: Agarose gel stained with ethidium bromide showing restriction digest analysis of plasmids used for barley transformation. All plasmids were digested with *Not*I and *Hin*dIII. Lane 1, pJH260; lane 2, pCO200 where the 4.2kb fragment corresponds to COL1A1 gene; lane 3, pCO210 that shows the 4.2 COL1A1 fragment and a 2.4kb fragment that corresponds to α subunit of P4H gene; lane 4 corresponds to pCO220, where can be seen the genes that code for COL1A1 (4.2kb) and β (2.3kb) subunits of P4H. The 3.0kb fragment observed in all the lanes correspond to a fragment from plasmid pJH260. Lane M, DNA marker

Table 3.2: *Agrobacterium tumefaciens* transformation.

Strain	Number of embryos	Green plants (SGM)	Green plants (RGM)	Plants in soil
AGL-1(COLA1)	100	15	0	3
AGL-1(COLA1+α)+ AGL-1(COLA1+β)†	600	31	14	4
AGL-1 (COLA1+ α+β)	200	N/A		

†: Co-cultivation with both strains

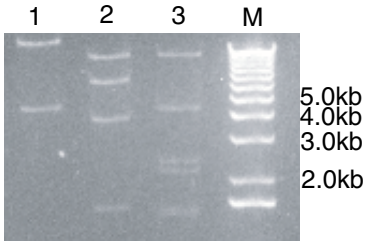

Figure 3.13: Agarose gel stained with ethidium bromide showing restriction digest analysis of plasmids pCO250 that contains the genes for hydroxylated procollagen expression in barley. Lane 1, digested with *Hin*dIII, where the 4.6kb fragment corresponds to α and β subunits genes; lane 2, digested with *Eco*RI, the 6.4kb fragment contains the genes COL1A1 and α subunit of P4H and the fragment with an apparent size of 4.0kb contains the β subunit gene and part of pJH260 plasmid; lane 3, digested with *Hin*dIII and *Eco*RI, showing the genes for COL1A1 (4.2kb), α (2.4kb) and β (2.3kb) subunits. The fragments of 1.5kb size in lanes 2 and 3 correspond to part of pJH260. Lane M, DNA marker

Chapter 4

Discussion

Transgenic plants represent a reproducible approach for high-level production of recombinant proteins. They can be grown, stored, processed and distributed with a relatively low level of technology. Moreover, protein secretion and folding in plants is similar to animals, which make them suitable for production of mammalian proteins. Based on these premises, the long term goal of this study is to produce recombinant collagen type I in the endosperm of barley grains, thereby generating an alternative to animal-derived collagen and minimizing possible immune responses to bone-extracted collagen. To achieve this goal, the gene was codon-optimized, assembled and successfully expressed in *Pichia pastoris*. In addition, barley transformants were obtained by transformation with *Agrobacterium tumefaciens*. At this time, transformants are being generated, and further analysis is needed to characterize the protein produced. It is known that hydroxylation is needed for the assembly of correctly folded and thermally stable collagen [*Bruckner and Prockop*, 1981*; Kivirikko and Pihlajaniemi*, 1998*; Fessler and Fessler*, 1978]. The question therefore is critical if barley is able to hydroxylate

the protein, and store it in protein bodies of the endosperm. Barley can synthesize up to 1g of recombinant protein per kilo of grain and store it in active form in the protein bodies of the endosperm [*Horvath et al.,* 2000]. In this study, the genes coding for alpha and beta subunits of prolyl 4-hydroxylase were co-transformed into immature zygotic embryos. It is expected that the genes will be transferred with two plasmids delivered by two different strains of *Agrobacterium* into the same cell as previously described [*Merle et al.,* 2002]. This will be a condition for the synthesis of hydroxylated and thermally stable homotrimeric collagen chains. Since the developing endosperm of barley contains PDI in its endoplasmic reticulum [*Mogelsvag and Simpson,* 1998] in the form of a homomeric dimer, it will be interesting to evaluate if the endogenous PDI can substitute for the β subunit of prolyl 4-hydroxylase in the transformants that only express the α subunit of prolyl 4-hydroxylase together with the gene encoding the α1 chain of type I collagen. This is of special interest since the prolyl 4-hydroxylase in tobacco cannot substitute for the human prolyl 4-hydroxylase [*Perret et al.,* 2001].

The expression of heterotrimeric collagen is also a point that needs to be consderated. Type I collagen is composed of two α1(I) and one α2(I) chains. Expression of heterotrimeric collagen in *Saccharomyces cerevisiae* [*Toman et al.,* 2000] was obtained using four genes. Thus, the next step for us would be to produce a plant with the four genes for production of collagen type I. These transformants can be made by at least three different approaches. The first one would be to make a new line that will have the COL2A1 gene and cross it with the existing transformants. This method is an

effective method for the assembly of complex proteins in plants [*Hiatt et al.,* 1989*; Ma et al.,* 1995]. An alternative method is to adapt the microspore transformation system recently developed for wheat [*Liu,* 2004] starting with a parental strain now produced. The main advantage of this method over use of immature embryos is the direct generation of transgenic homozygous doubled haploid lines in 6 months. This has provided the possibility for a faster production of transgenic cereals. These two methods have the challenge that the selection of the transformants has to be done with a different selectable marker.

One possible alternative to bialaphos selection is the use of the phosphomannose isomerase gene, *pmi* (*manA* from *Escherichia coli*), as a selectable marker. Phosphomannose isomerase (PMI) converts mannose-6-phosphate, an unmetabolizable carbon source for most plant cells, into fructose-6-phosphate, a carbohydrate source that can be used by plants. Plants expressing transgenic *pmi* are able to metabolize mannose as a carbon source. This promotes growth of transgenic tissues on media containing mannose, while non-transformed calli or plants either stop growing or die due to starvation [*Wright et al.,* 2001*; Hansen and Wright,* 1999]. Moreover, the number of escapes produced using this selection method decreases considerably, by generating two to three fold more transformants than obtained with *bar* gene as the selectable marker [*Wright et al.,* 2001]. Successful results have been reported on wheat, corn [*Wright et al.,* 2001] and rice [*Datta et al.,* 2003].

The biosynthesis of collagen requires lysyl hydroxylase, also named procollagen-

lysine, 2-oxoglutarate 5-dioxygenase (PLOD). Lysyl hydroxylase is an important post-translational modifying enzyme in collagen biosynthesis [*Kivirikko and Pihlajaniemi,* 1998]. This enzyme hydroxylates specific lysine residues in the collagen molecule to form hydroxylysines that have two important functions: attachment sites for galactose and glucosylgalactose and precursors for the cross-linking process that gives collagen its tensile strength [*Kivirikko and Pihlajaniemi,* 1998*; Notbohm et al.,* 1999*; Kivirikko et al.,* 1990]. The number of hydroxylated lysyl residues and glycosylated hydroxylysine residues varies not only among different collagen types but also within the same collagen type in different tissues and under different physiological conditions. Embryonic collagens, for instance, are more extensively modified than adult collagens [*Prockop and Kivirikko,* 1995].

The importance of lysyl hydroxylase was demonstrated by in vitro fibril formation of collagen type II, using a bacilovirus system. Fully processed fibrils were formed after infection with the vector coding for lysil hydroxylase [*Notbohm et al.,* 1999]. Three isoforms (LH1, LH2, LH3) have been characterized from human and mouse. [*Kivirikko et al.,* 1990][*Valtavaara and an R. Myllyla,* 1998]. In later studies, LH3 was shown to be a multifunctional protein possessing both collagen galactosyl transferase and collagen glucosyltransferase activity [*Wang et al.,* 2002] even though the levels of the glycosyltransferase activities may not have been biologically significant [*Rautavuoma et al.,* 2002].

The results presented here are preliminary and were designed to evaluate the po-

tential of using barley as a producer of recombinant collagen. The developing transformants have to be characterized and their usefulness evaluated. It is then of high priority to include the gene that codes for lysyl hydroxylase in the production of transgenic plants by transformation of zygotic barley embryos. Alternatively, the potential for transformation of barley microspores for a faster generation of homozygous transgenic plants by chromosome doubling during regeneration of the developing microspores into mature plants should be evaluated.

Bibliography

Bassuk, J. A., and R. A. Berg, Protein disulphide isomerase, a multifunctional endoplasmic reticulum protein, *Matrix*, *9*, 244–258, 1989.

Batard, Y., A. Hehn, S. Nedelkina, M. Schalk, K. Pallett, H. Schaller, and D. Werck-Reichhart, Increasing expression of P450 and P450-reductase proteins from monocots in heterologous systems, *Archives of Biochemistry and Biophysics*, *379*, 161–169, 2000.

Bateman, J. F., S. R. Lamande, and J. M. Ramshaw, Collagen superfamily, in *Extracellular matrix. Molecular components and interactions*, edited by W. Comper, pp. 22–67, Harwood Academic Publishers, 1996.

Bornstein, P., The biosynthesis, secretion and processing of procollagen, in *Biology of Collagen*, edited by A. Viidik, and J. Vuust, pp. 61–75, New York, Academic Press, London, 1980.

Brodsky, B., and J. A. M. Ramshaw, The collagen triple-helix structure, *Matrix Biology*, *15*, 545–554, 1997.

Bruckner, P., and D. J. Prockop, Proteolytic enzymes as probes for the triple-helical conformation of procollagen, *Analytical Biochemistry*, *110*, 360–368, 1981.

Bulleid, N. J., D. C. John, and K. E. Kadler, Recombinant expression systems for the production of collagen, *Biochemical Society Transactions*, *28*, 350–353, 2000.

Cameron-Mills, V., and D. von Wettstein, Protein body formation in the developing barley endosperm, *Carlsberg Res. Commun.*, *45*, 577–594, 1980.

Cameron-Mills, V., J. Ingversen, and A. Brandt, Transfer of in vitro synthesized barley endosperm poteins into the lumen of the endoplasmic reticulum, *Carlsberg Res. Commun.*, *43*, 91–102, 1978.

Chiapello, H., F. Lisacek, M. Caboche, and A. Henaut, Codon usage and gene function are related in sequences of *Arabidopsis thaliana*, *Gene*, *209*, GC1–GC38, 1998.

Cho, M., H. Choi, W. Jiang, C. Ha, and P. Lemaux, Endosperm-specific expression of green fluorescent protein driven by the hordein promoter is stably inherited in transgenic barley *Hordeum vulgare* plants, *Physiologia Plantarum*, *115*, 144–154, 2002.

Cregg, J., K. Barringer, A. Hessler, and K. Madden, *Pichia pastoris* as a host system for transformations, *Mollecular and Cell Biology*, *5*, 3376–3385, 1985.

Datta, K., N. Baisakh, N. Oliva, L. Torrizo, E. Abrigo, J. Tang, M. Rai, S. Rehana, S. Al-Babili, P. Beyer, I. Potrykus, and S. K. Datta, Bioengineered 'golden' indica

rice cultivars with β-carotene metabolism in the endosperm with hygromycin and mannose selection systems, *Plant Biotechnology Journal*, *1*, 81–90, 2003.

De Block, M., A. D. Sonville, and D. Debrouwer, The selection mechanism of phosphinothricin is influenced by the metabolic status of the tissue, *Planta*, *197*, 619–626, 1995.

de Bruin, E. C., F. A. de Wolf, and N. C. M. Laane, Expression and secretion of human α1(I) procollagen fragment by *Hansenula polymorpha* as compared to *Pichia pastoris*, *Enzyme and Microbial Technology*, *26*, 640–644, 2000.

de Bruin, E. C., M. W. T. Werten, C. Laane, and F. A. de Wolf, Endogenous prolyl 4-hydroxylation in *Hansenula polymorpha* and its use for the production of hydroxylated recombinant gelatin, *FEMS Yeast Research*, *1*, 291–298, 2002.

Engel, J., and D. J. Prockop, The zipper-like folding of collagen triple helices and the effects of mutations that disrupt the zipper, *Annual Review of Biophysical Chemistry*, *20*, 137–152, 1991.

Fessler, J. H., and L. I. Fessler, Biosynthesis of collagens, *Annual Review of Biochemistry*, *47*, 129–162, 1978.

Fichard, A., E. Tillet, F. Delacoux, R. Garrone, and F. Ruggiero, Human recombinant a1(V) collagen chain: Homotrimeric assembly and subsequent processing, *The Journal of Biological Chemistry*, *272*, 30083–30087, 1997.

Fischer, R., R. Twyman, and S. Schillberg, Production of antibodies in plants and their use for global health, *Vaccine*, *21*, 820–825, 2003.

Friess, W., Collagen- biomaterial for drug delivery, *European Journal of Pharmaceutics and Biopharmaceutics*, *45*, 113–136, 1998.

Hansen, G., and M. Wright, Recent advances in the transformation of plants, *Trends in Plant Science*, *4*, 226–231, 1999.

Hiatt, A., R. Cafferkey, and K. Bowdish, Production of antibodies in transgenic plants, *Nature*, *342*, 76–78, 1989.

Hood, E., From green plants to industrial enzymes, *Enzyme and Microbiology Technology*, *30*, 279–83, 2002.

Horvath, H., J. T. Huang, O. T. Wong, and D. von Wettstein, The production of recombinant proteins in transgenic barley grains, *Proceedings of the National Academy of Sciences USA*, *97*, 1914–1917, 2000.

Horvath, H., L. Jensen, O. Wong, E. Kohl, S. Ullrich, J. Cochran, C. Kannangara, and D. von Wettstein, Stability of transgene expression, field performance and recombination breeding of transformed barley lines, *Theor. Appl. Genet.*, *102*, 1–11, 2001.

John, D. C., R. Watson, A. Kind, A. R. Scott, K. E. Kadler, and N. J. Bulleid, Expression of an engineered form of recombinant procollagen in mouse milk, *Nature Biotechnology*, *17*, 385–389, 1999.

Kadler, K., D. F. Holmes, J. A. Trotter, and J. A. Chapman, Collagen fibril formation, *Biochem. J.*, *316*, 1–11, 1996.

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

Kapusta, J., A. Modelska, M. Figlerowicz, T. Pniewski, M. Letellier, O. Lisowa, V. Yusibov, H. Koprowski, A. Plucienniczak, and A. Legocki, A plant- derived edible vaccine against hepatitis B virus, *FASEB J.*, *13*, 1796–1799, 1999.

Keizer-Gunnink, I., A. Vuorela, J. Myllyharju, T. Pihlajaniemi, K. Kivirikko, and M. Veenhuis, Accumulation of properly folded human type III procollagen molecules in specific intracellular membranous compartments in the yeast *Pichia pastoris*, *Matrix Biology*, *19*, 29–36, 2000.

Kivirikko, K. I., Post-translational modifications of collagen, in *Gene families of collagen and other proteins*, edited by D. J. Prockop, and Champe, pp. 107–119, Elsevier North Holland Inc., 1980.

Kivirikko, K. I., Collagens and their abnormalities in a wide spectrum of diseases, *Annals of Medicine*, *25*, 113–126, 1993.

Kivirikko, K. I., Collagen biosynthesis: a mini-review cluster, *Matrix Biolgy*, *16*, 355–356, 1998.

Kivirikko, K. I., and T. Pihlajaniemi, Collagen hydroxylases and the protein disulfide isomerase subunit of prolyl 4-hydroxylases, *Advances in Enzymology and Related Areas of Molecular Biology*, *72*, 325–398, 1998.

Kivirikko, K. I., R. Myllyla, and T. Pihlajaniemi, Hydroxylation of proline and lysine residues in collagen and other animal and plant proteins, in *Post-translational modifications of proteins*, edited by J. J. Harding, and M. J. C. Crabbe, pp. 1–52, CRC Press, 1990.

Koziel, M., N. Carozzi, and N. Desai, Optimizing expression of transgenes with an emphasis on post-transcriptional events, *Plant Molecular Biology*, *32*, 393–405, 1996.

Leibovich, S. J., and J. B. Weiss, Electron microscope studies of the effects of endo- and exopeptidase digestion on tropocollagen. A novel concept of the role of terminal regions in fibrillogenesis., *Biochimical and Biophysical Acta*, *214*, 445–454, 1970.

Linder, S., M. Schliwa, and E. Kube-Granderath, Direct PCR screening of *Pichia pastoris* clones, *BioTechniques*, *20*, 980–982, 1996.

Liu, W., *Transformation of microspores for generating double haploid transgenic wheat (Triticum aestivum L.)*, Ph.D Thesis. Washington State University, Pullman, WA, 2004.

Lynn, A. K., I. V. Yannas, and W. Bonfield, Antigenicity and immunogenicity of collagen, *Journal of Biomedical Materials Research Part B: Applied Biomaterials*, *71B*, 343–354, 2004.

Ma, J., A. Hiatt, M. Hein, N. Vine, F. Wang, P. Stabila, C. Dolleweerd, K. Mostov, and T. Lehner, Generation and assembly of secretory antibodies in plants, *Science*, *268*, 716–719, 1995.

Ma, J., P. M. Drake, and P. Christou, The production of recombinant pharmaceutical proteins in plants, *Nature Reviews/ Genetics*, *4*, 794–805, 2003.

McLaughlin, S. H., and N. J. Bulleid, Molecular recognition in procollagen chain assembly, *Matrix Biology*, *16*, 369–377, 1998.

Merle, C., S. Perret, T. Lacour, V. Jonval, S. Hudaverdian, R. Garrone, F. Ruggiero, and M. Theissen, Hydroxylated human homotrimeric collagen I in *Agrobacterium tumefaciens*- mediated transient expression and in transgenic tobacco plant, *FEBS Letters*, *515*, 114–118, 2002.

Mogelsvag, S., and D. Simpson, Protein folding and transport from the endoplasmic reticulum to the Golgi apparatus in plants, *Journal of Plant Physiology*, *153*, 1–15, 1998.

Murray, E., J. Lotzer, and M. Eberle, Codon usage in plant genes, *Nucleic Acids Research*, *17*, 477–498, 1989.

Myllyharju, J., Prolyl 4-hydroxylases, the key enzymes of collagen synthesis, *Matrix Biology*, *22*, 15–24, 2003.

Myllyharju, J., A. Lamberg, H. Notbohm, P. Fietzek, T. Pihlajaniemi, and K. I. Kivirikko, Expression of wild-type and modified proα chains of human type I procollagen in insect cells leads to the formation of stable [α1(I)]2α2(I) collagen heterotrimers and [α1(I)]3 homotrimers but not [α2(I)]3 homotrimers, *The Journal of Biological Chemistry*, *272*, 21824–21830, 1997.

Myllyharju, J., and K. I. Kivirikko, Collagens and collagen-related diseases, *Annals of Medicine*, *33*, 7–21, 2001.

Myllyharju, J., M. Nokelainen, A. Vuorela, and K. I. Kivirikko, Expression of recombinant human type I-III collagen in the yeast *Pichia pastoris*, *Biochemical Society Transactions*, *28*, 353–357, 2000.

Nauerby, B., K. Billing, and R. Wyndaele, Influence of the antibiotic timentin on plant regeneration compared to carbenicillin and cefotaxime in concentrations suitable for elimination of *Agrobacterium tumefaciens*, *Plant Science*, *123*, 169–177, 1997.

Nokelainen, M., H. Tu, A. Vuorela, H. Notbohm, K. I. Kivirikko, and J. Myllyharju, High-level production of human type I collagen in the yeast *Pichia pastoris*, *Yeast*, *18*, 797–806, 2001.

Notbohm, H., M. Nokelainen, J. Myllyharju, P. P. Fietzek, P. K. Muller, and K. I. Kivirikko, Recombinant human type II collagens with low and high levels of hydroxylysine and its glycosylated forms show marked differences in fibrillogenesis in vitro, *The Journal of Biological Chemistry*, *274*, 8988–8992, 1999.

Olsen, D., C. Yang, M. Bodo, R. Chang, S. Leigh, J. Baez, D. Carmichael, M. Perala, E. Hamalainen, M. Jarvinen, and J. Polarek, Recombinant collagen and gelatin for drug delivery, *Advanced Drug Delivery Reviews*, *55*, 1547–1567, 2003.

Pakkanen, O., E. Hamalainen, K. Kivirikko, and J. Myllyharju, Assembly of stable human type i and III collagen molecules from hydroxylated recombinant chains in the yeast *Pichia pastoris*, *The Journal of Biological Chemistry*, *278*, 32478–32483, 2003.

Paleyanda, R., W. Velander, T. Lee, D. Scandella, F. Gwazdauskas, J. Knight, L. Hoyer, W. Drohan, and H. Lubon, Transgenic pigs produce functional human factor VIII in milk, *Nature Biotechnology*, *15*, 971–975, 1997.

Perret, S., C. Merle, S. Bernocco, P. B. an R. Garrone, D. J. S. Hulmes, M. Theisen, and F. Ruggiero, Unhydroxylated triple helical collagen I produced in transgenic plants provides new clues on the role of hydroxyproline in collagen folding and fibril formation, *The Journal of Biological Chemistry*, *276*, 43693–43698, 2001.

Peterson, R. K., and C. J. Arntzen, On risk and plant- based biopharmaceuticals, *TRENDS in Biotechnology*, *22*, 64–66, 2004.

Pihlajaniemi, T., T. Helaakosky, K. Tasanen, R. Myllyla, M. Huhtala, J. Koivu, and K. I. Kivirikko, Molecular cloning of the β subunit of human prolyl 4-hydroxylase. This subunit and protein disulfide isomerase are products of the same gene, *The EMBO Journal*, *6*, 643–649, 1987.

Prockop, D. J., A. L. Sieron, and S.-W. Li, Procollagen N-proteinase and procollagen C-proteinase. two unusual metalloproteinases that are essential for procollagen processing probably have important roles in development and cell signaling, *Matrix Biology*, *16*, 399–408, 1998.

Prockop, D. J., and K. I. Kivirikko, Collagens: molecular biology, diseases, and potentials for therapy, *Annual Review of Biochemistry*, *64*, 403–434, 1995.

Prunkard, D., I. Cottingham, I. Garner, S. Bruce, M. Dalrymple, G. Lasser, P. Bishop, and D.Foster, High-level expression of recombinant human fibrinogen in the milk of transgenic mice, *Nature Biotechnology*, *14*, 867–861, 1996.

Rautavuoma, K., K. Takaluoma, K. Passoja, A. Pirskanen, A. P. Kvist, K. I. Kivirikko, and J. Myllyharju, Characterization of three fragments that constitute the monomers of the human lysyl hydroxylase isoenzymes 1-3. The 30-kDa N-terminal fragment is not required for lysyl hydroxylase activity, *The Journal of Biological Chemistry*, *277*, 23084–23091, 2002.

Ruggiero, F., J. Y. Exposito, P. Bournat, V. Gruber, S. Perret, J. Comte, B. Olagnier, R. Garrone, and M. Theisen, Triple helix assembly and processing of human

collagen produced in transgenic tobacco plants, *FEBS Letters*, *469*, 132–136, 2000.

Sambrook, S., E. F. Fritsch, and T. Maniatis, *Molecular cloning: A Laboratory Manual*, second ed., Cold Spring Harbor Laboratory Press, Plainview, N. Y., 1989.

Sijmons, P. C., B. M. Dekker, B. Schrammeijer, T. C. Verwoerd, P. J. van den Elzen, and A. Hoekema, Production of correctly processed human serum albumin in transgenic plants, *Biotechnology*, *8*, 217–221, 1990.

Stahl, R., H. Horvath, J. van Fleet, M. Voetz, D. von Wettstein, and N. Wolf, T-DNA integration into barley genome from single and double cassette vectors, *Proceedings of the National Academy of Sciences USA*, *99*, 2146–2151, 2002.

Streatfield, S., J. Lane, C. Brooks, D. Barker, M. Poage, J. Mayor, B. Lamphear, C. Drees, J. Jilka, E. Hood, and J. Howard, Corn as production system for human and animal vaccines, *Vaccine*, *21*, 812–815, 2003.

Stromqvist, M., M. Houdebine, J. Andersson, A. Edlund, T. Johansson, C. Viglietta, C. Puissant, and L. Hansson, Recombinant human extracellular superoxide dismutase produced in milk of transgenic rabbits, *Transgenic Research*, *6*, 271–278, 1997.

Tanaka, M., K. Sato, and T. Uchida, Plant prolyl hydroxylase recognizes poly(l-proline) II helix, *The Journal of Biological Chemistry*, *256*, 11397–11400, 1981.

Tingay, S., D. McElroy, R. Kalla, S. Fieg, M. Wang, S. Thorton, and R. Bretell, *Agrobacterium tumefaciens* barley transformation, *The Plant Journal*, *11*, 1369–1376, 1997.

Toman, P. D., F. Pieper, N. Sakai, C. Karatzas, E. Platenburg, I. de Wit, A. Dekker, G. Daniels, R. Berg, and G. Platenburg, Production of recombinant human type I procollagen homotrimers in the mammary gland of transgenic mice, *Transgenic Research*, *8*, 415–427, 1999.

Toman, P. D., G. Chisholm, H. McMullin, L. M. Giere, D. R. Olsen, R. J. Kovach, S. D. Leigh, B. E. Fong, R. C. an G. A. Daniels, R. A. Berg, and R. A. Hitzeman, Production of recombinant human type I procollagen trimers using a four-gene expression system in the yeast *Saccharomyces cerevisiae*, *The Journal of Biological Chemistry*, *275*, 23303–23309, 2000.

Tomita, M., H. Munetsuna, T. Sato, T. Adachi, R. Hino, M. Hayashi, K. Shimizu, N. N. an T. Tamura, and K. Yoshizato, Transgenic silkwoms produce recombinant human type III procollagen in cocoons, *Nature Biotechnology*, *21*, 52–56, 2003.

Twyman, R. M., E. Stoger, S. Schillberg, P. Christou, and R. Fischer, Molecular farming in plants: host systems and expression technology, *TRENDS in Biotechnology*, *21*, 570–578, 2003.

Valtavaara, M. C. S., and J. S. an R. Myllyla, Primary structure, tissue distribution, and chromosomal localization of a novel isoform of lysyl hydroxylase (lysyl

hydroxylase 3), *The Journal of Biological Chemistry*, *273*, 12881–12886, 1998.

van der Rest, M., and R. Garrone, Collagen family of proteins, *FASEB J.*, *5*, 2814–2823, 1991.

van Fleet, J., *Production of Recombinant Human Lysozyme and Lactoferrin in Transgenic Barley Grains*, M.S. Thesis. Washington State University, Pullman, WA, 2001.

Vranka, J., L. Sakai, and H. Bachinger, Prolyl 3-hydroxylase 1, enzyme characterization and identification of a novel family of enzymes, *The Journal of Biological Chemistry*, *279*, 23615–23621, 2004.

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

Wang, C., H. Luosujarvi, J. Heikkinen, M. Risteli, and L. U. an R. Myllyla, The third activity for lysyl hydroxylase 3: galactosylation of hydroxyl residues in collagens in vitro, *Matrix Biology*, *21*, 559–566, 2002.

Westerhausen, A., C. D. Constantinou, M. Pack, M. Peng, C. Hanning, A. S. Olsen, and D. J. Prockop, Completion of the last half of the structure of the human gene for the Proa1(I) chain of type i procollagen (col1a1), *Matrix*, *11*, 375–379, 1991.

Wilson, R., J. F. Lees, and N. J. Bulleid, Protein disulfide isomerase acts as a molecular chaperone during tha assembly of procollagen, *The Journal of Biological Chemistry*, *273*, 9637–9643, 1998.

Wright, M., J. Dawson, E. Dunder, J. Suttie, J. Reed, C. Kramer, Y. Chang, and H. Wang, Efficient biolistic transformation of maize (*Zea mays* l.) and wheat (*Triticum aestivum* l.) using the phosphomannose isomerase, *pmi*, gene as the selectable marker, *Plant Cell Reproduction*, *20*, 429–436, 2001.

Wu, Y., *Transformation of Barley for Resistance to Rhizoctonia Root Rot*, Ph.D Thesis. Washington State University, Pullman, WA, 2003.

Chapter 5

Appendices

5.1 Oligos for gene assembly

Col 1(99 nt): CCG CTC AGC TGA GCT ACG GCT ACG ACG AGA AGA GCA CCG GAG GTA TCA GCG TGC CTG GCC GCA TGG GTC CGA GCG GTC CAA GGG GAC TGC CTG GCC CAC

Col 2 (99 nt): CTG GTG CTC CTG GAC CTC AGG GAT TTC AAG GAC CAC CTG GAG AAC CTG GAG AGC CGG GAG CCT CTG GAC CTA TGG GCC CAA GGG GAC CTC CGG GAC CAC

Col 3 (99 nt): CTG GTA AGA ATG GAG ACG ACG GCG AGG CTG GTA AGC CCG GGA GGC CAG GAG AGA GGG GAC CAC CAG GAC CGC AGG GCG CTA GGG GTC TGC CGG GGA CAG

Col 4 (99 nt): CTG GAC TGC CAG GCA TGA AGG GAC ACA GGG GTT TCA GCG GTC TAG ACG GAG CTA AGG GGG ACG CTG GAC CAG CAG GAC CCA AGG GTG AGC CAG GAT CTC

Col 5 (98 nt): CAG GAG AAA ACG GCG CGC CAG GTC AGA TGG GAC CAA GAG GCC TGC CCG GTG AGA GAG GTA GAC CAG GAG CGC CCG GTC CAG CTG GTG CCA GGG GAA AC

Col 6 (99 nt): GAT GGT GCT ACA GGA GCG GCC GGT CCA CCT GGT CCT ACT GGT CCC GCC GGT CCT CCT GGA TTC CCT GGT GCC GTT GGA GCT AAG GGT GAG GCA GGT CCG

Col 7 (99 nt): CAG GGG CCA AGG GGT AGC GAA GGA CCT CAA GGA GTG CGT GGT GAG CCT GGG CCG CCG GGT CCT GCT GGT GCC GCT GGT CCC GCT GGA AAC CCA GGT GCC

Col 8 (100 nt): GAC GGT CAA CCA GGA GCC AAA GGC GCC AAC GGT GCA CCA GGG ATC GCA GGA GCC CCA GGC TTT CCA GGA GCT AGA GGC CCA AGC GGA CCT CAA GGA CCT G

Col 9 (100 nt): GTG GCC CAC CTG GAC CGA AGG GTA ACT CTG GAG AGC CCG GAG CCC CAG GAA GCA AAG GTG ACA CTG GAG CCA AGG GTG AGC CTG GAC CGG TTG GTG TAC A

Col 10 (100 nt): GGG ACC GCC AGG ACC AGC CGG TGA GGA GGG AAA GAG GGG CGC TAG GGG TGA GCC TGG ACC AAC TGG ACT GCC TGG ACC ACC TGG TGA GAG GGG CGG CCC T

Col 11 (100 nt): GGT AGC AGA GGA TTC CCT GGC GCT GAC GGA GTT GCT GGA CCT AAG GGA CCA GCT GGA GAG AGG GGA TCA CCA GGA CCT GCC GGA CCG AAG GGA TCT CCA G

Col 12 (99 nt): GCG AAG CAG GTA GGC CAG GTG AAG CAG GAC TGC CAG GTG CCA AAG GAC TGA CAG GCT CCC CTG GAT CTC CTG GTC CTG ACG GTA AGA CTG GCC CTC CTG

Col 13 (100 nt): GAC CTG CTG GTC AAG ATG GGA GAC CTG GAC CAC CGG GAC CAC CTG GAG CTA GGG GAC AAG CTG GCG TGA TGG GTT TTC CTG GGC CAA AGG GAG CTG CAG G

Col 14 (99 nt): CGA ACC TGG TAA GGC TGG CGA GAG GGG AGT TCC AGG TCC TCC AGG TGC CGT GGG TCC CGC TGG AAA GGA TGG TGA GGC AGG TGC ACA GGG TCC GCC AGG

Col 15 (99 nt): GCC TGC TGG TCC AGC CGG TGA GAG GGG GGA GCA AGG ACC TGC CGG ATC GCC AGG TTT CCA GGG ACT GCC GGG ACC TGC TGG GCC ACC TGG TGA AGC TGG

Col 16 (100 nt): GAA ACC GGG CGA GCA GGG CGT GCC AGG AGA TCT AGG GGC TCC TGG GCC AAG CGG TGC TAG GGG TGA GAG GGG CTT TCC AGG AGA GAG AGG AGT GCA AGG A

Col 17 (100 nt): CCA CCT GGG CCG GCT GGA CCT AGA GGC GCT AAC GGA GCA CCA GGT AAC GAT GGA GCT AAG GGA GAC GCA GGC GCA CCT GGA GCA CCG GGA TCA CAG GGA G

Col 18 (100 nt): CAC CAG GAC TGC AGG GCA TGC CAG GTG AGC GTG GAG CTG CGG GCC TGC CTG GTC CCA AGG GAG ACC GCG GCG ACG CTG GTC CTA AAG GTG CGG ACG GAA G

Col 19 (98 nt): CCC TGG CAA GGA CGG AGT GAG AGG TCT GAC TGG CCC TAT CGG TCC TCC TGG TCC AGC TGG CGC GCC CGG TGA CAA AGG TGA GAG CGG CCC ATC TGG TC

Col 20 (99 nt): CTG CAG GTC CGA CTG GTG CCA GGG GGG CTC CCG GCG ACA GAG GTG AGC CAG GCC CTC CTG GTC CAG CTG GTT TCG CGG GAC CTC CAG GTG CCG ACG GTC

Col 21 (99 nt): AGC CAG GCG CAA AGG GAG AGC CCG GTG ACG CAG GAG CGA AGG GAG ATG CAG GGC CAC CAG GAC CAG CGG GAC CGG CCG GAC CAC CTG GAC CAA TCG GTA

Col 22 (100 nt): ACG TGG GTG CAC CTG GGG CTA AGG GCG CTA GGG GTT CTG CAG GTC CTC CTG GAG CCA CTG GTT TCC CTG GAG CCG CCG GTA GAG TTG GAC CAC CTG GAC C

Col 23 (100 nt): GTC TGG AAA CGC AGG ACC ACC GGG ACC ACC TGG GCC AGC GGG AAA GGA AGG AGG CAA AGG GCC AAG AGG CGA GAC TGG ACC AGC AGG ACG TCC AGG TGA G

Col 24 (100 nt): GTT GGA CCT CCA GGA CCC CCA GGC CCA GCA GGA GAG AAA GGT AGC CCA GGT GCA GAT GGC CCA GCT GGC GCG CCC GGT ACT CCA GGC CCA CAG GGT ATT G

Col 25 (100 nt): CAG GAC AGA GGG GCG TGG TGG GTC TGC CAG GAC AGA GGG GGG AGA GGG GTT TTC CAG GCC TGC CGG GTC CTT CTG GGG AGC CAG GAA AGC AGG GAC CTA G

Col 26 (100 nt): CGG TGC CAG CGG AGA GAG GGG GCC ACC TGG TCC GAT GGG TCC TCC GGG GCT AGC TGG TCC ACC TGG AGA GTC TGG TAG GGA GGG TGC ACC GGG CGC CGA A

Col 27 (100 nt): GGC TCA CCA GGA CGT GAT GGT TCG CCA GGT GCC AAA GGG GAT AGG GGA GAG ACA GGA CCG GCA GGA CCA CCT GGT GCT CCA GGC GCC CCG GGG GCT CCA G

Col 28 (100 nt): GAC CTG TCG GTC CAG CTG GAA AGT CAG GTG ACA GAG GAG AGA CTG GCC CAG CAG GAC CTG CGG GAC CGG TGG GAC CAG TGG GTG CCA GGG GAC CAG CAG G

Col 29 (99 nt): GCC TCA GGG ACC GCG TGG AGA CAA GGG TGA GAC CGG AGA GCA GGG CGA CAG GGG TAT CAA GGG GCA CAG GGG GTT CAG CGG TCT GCA GGG CCC TCC AGG

Col 30 (100 nt): ACC ACC TGG TTC ACC GGG CGA GCA AGG ACC ATC AGG CGC AAG CGG ACC AGC AGG GCC TCG CGG ACC TCC AGG ATC TGC CGG CGC CCC AGG TAA GGA CGG T

Col 31 (100 nt): CTG AAT GGT CTC CCA GGA CCT ATT GGA CCG CCA GGG CCT AGG GGT CGT ACG GGT GAC GCT GGA CCT GTG GGC CCG CCG GGA CCA CCA GGA CCA CCA GGA C

Col 32 (99 nt): CTC CAG GCC CTC CAA GCG CAG GTT TCG ACT TCA GCT TTC TGC CAC AAC CTC CAC AGG AGA AGG CCC ACG ACG GTG GAA GGT ACT ACA GAG CCT ACA TCC

Col 33 (93 nt): CCG AAG CCC CGC GCG ATG GTC AGG CCT ACG TGA GAA AGG ACG GCG AGT GGG TCT TCC TGA GCA CCT TCC TGA GCC CTG CCT GAG AGC TCG CGC

Col 34R (52 nt): GCG CGA GCT CTC AGG CAG GGC TCA GGA AGG TGC TCA GGA AGA CCC ACT CGC C

Col 35R (99 nt): GTC CTT TCT CAC GTA GGC CTG ACC ATC GCG CGG GGC TTC GGG GAT GTA GGC TCT GTA GTA CCT TCC ACC GTC GTG GGC CTT CTC CTG TGG AGG TTG TGG

Col 36R (99 nt): CAG AAA GCT GAA GTC GAA ACC TGC GCT TGG AGG GCC TGG AGG TCC TGG TGG TCC TGG TGG TCC CGG CGG GCC CAC AGG TCC AGC GTC ACC CGT ACG ACC

Col 37R (99 nt): CCT AGG CCC TGG CGG TCC AAT AGG TCC TGG GAG ACC ATT CAG ACC GTC CTT ACC TGG GGC GCC GGC AGA TCC TGG AGG TCC GCG AGG CCC TGC TGG TCC

Col 38R (98 nt): GCT TGC GCC TGA TGG TCC TTG CTC GCC CGG TGA ACC AGG TGG TCC TGG AGG GCC CTG CAG ACC GCT GAA CCC CCT GTG CCC CTT GAT ACC CCT GTC GC

Col 39R (98 nt): CCT GCT CTC CGG TCT CAC CCT TGT CTC CAC GCG GTC CCT GAG GCC CTG CTG GTC CCC TGG CAC CCA CTG GTC CCA CCG GTC CCG CAG GTC CTG CTG GG

Col 40R (100 nt): CCA GTC TCT CCT CTG TCA CCT GAC TTT CCA GCT GGA CCG ACA GGT CCT GGA GCC CCC GGG GCG CCT GGA GCA CCA GGT GGT CCT GCC GGT CCT GTC TCT C

Col 41R (100 nt): CCC TAT CCC CTT TGG CAC CTG GCG AAC CAT CAC GTC CTG GTG AGC CTT CGG CGC CCG GTG CAC CCT CCC TAC CAG ACT CTC CAG GTG GAC CAG CTA GCC C

Col 42R (99 nt): CGG AGG ACC CAT CGG ACC AGG TGG CCC CCT CTC TCC GCT GGC ACC GCT AGG TCC CTG CTT TCC TGG CTC CCC AGA AGG ACC CGG CAG GCC TGG AAA ACC

Col 43R (100 nt): CTC TCC CCC CTC TGT CCT GGC AGA CCC ACC ACG CCC CTC TGT CCT GCA ATA CCC TGT GGG CCT GGA GTA CCG GGC GCG CCA GCT GGG CCA TCT GCA CCT G

Col 44R (100 nt): GGC TAC CTT TCT CTC CTG CTG GGC CTG GGG GTC CTG GAG GTC CAA CCT CAC CTG GAC GTC CTG CTG GTC CAG TCT CGC CTC TTG GCC CTT TGC CTC CTT C

Col 45R (98 nt): CTT TCC CGC TGG CCC AGG TGG TCC CGG TGG TCC TGC GTT TCC AGA CGG TCC AGG TGG TCC AAC TCT ACC GGC GGC TCC AGG GAA ACC AGT GGC TCC AG

Col 46R (99 nt): GAG GAC CTG CAG AAC CCC TAG CGC CCT TAG CCC CAG GTG CAC CCA CGT TAC CGA TTG GTC CAG GTG GTC CGG CCG GTC CCG CTG GTC CTG GTG GCC CTG

Col 47R (99 nt): CAT CTC CCT TCG CTC CTG CGT CAC CGG GCT CTC CCT TTG CGC CTG GCT GAC CGT CGG CAC CTG GAG GTC CCG CGA AAC CAG CTG GAC CAG GAG GGC CTG

Col 48R (99 nt): GCT CAC CTC TGT CGC CGG GAG CCC CCC TGG CAC CAG TCG GAC CTG CAG GAC CAG ATG GGC CGC TCT CAC CTT TGT CAC CGG GCG CGC CAG CTG GAC CAG

Col 49R (100 nt): GAG GAC CGA TAG GGC CAG TCA GAC CTC TCA CTC CGT CCT TGC CAG GGC TTC CGT CCG CAC CTT TAG GAC CAG CGT CGC CGC GGT CTC CCT TGG GAC CAG G

Col 50R (99 nt): CAG GCC CGC AGC TCC ACG CTC ACC TGG CAT GCC CTG CAG TCC TGG TGC TCC CTG TGA TCC CGG TGC TCC AGG TGC GCC TGC GTC TCC CTT AGC TCC ATC

Col 51R (99 nt): GTT ACC TGG TGC TCC GTT AGC GC CTC TAG GTC CAG CCG GCC CAG GTG GTC CTT GCA CTC CTC TCT CTC CTG GAA AGC CCC TCT CAC CCC TAG CAC CGC T

Col 52R (99 nt): TGG CCC AGG AGC CCC TAG ATC TCC TGG CAC GCC CTG CTC GCC CGG TTT CCC AGC TTC ACC AGG TGG CCC AGC AGG TCC CGG CAG TCC CTG GAA ACC TGG

Col 53R (100 nt): CGA TCC GGC AGG TCC TTG CTC CCC CCT CTC ACC GGC TGG ACC AGC AGG CCC TGG CGG ACC CTG TGC ACC TGC CTC ACC ATC CTT TCC AGC GGG ACC CAC G

Col 54R (100 nt): GCA CCT GGA GGA CCT GGA ACT CCC CTC TCG CCA GCC TTA CCA GGT TCG CCT GCA GCT CCC TTT GGC CCA GGA AAA CCC ATC ACG CCA GCT TGT CCC CTA G

Col 55R (100 nt): CTC CAG GTG GTC CCG GTG GTC CAG GTC TCC CAT CTT GAC CAG CAG GTC CAG GAG GGC CAG TCT TAC CGT CAG GAC CAG GAG ATC CAG GGG AGC CTG TCA G

Col 56R (98 nt): TCC TTT GGC ACC TGG CAG TCC TGC TTC ACC TGG CCT ACC TGC TTC GCC TGG AGA TCC CTT CGG TCC GGC AGG TCC TGG TGA TCC CCT CTC TCC AGC TG

Col 57R (100 nt): GTC CCT TAG GTC CAG CAA CTC CGT CAG CGC CAG GGA ATC CTC TGC TAC CAG GGC CGC CCC TCT CAC CAG GTG GTC CAG GCA GTC CAG TTG GTC CAG GCT C

Col 58R (97 nt): ACC CCT AGC GCC CCT CTT TCC CTC CTC ACC GGC TGG TCC TGG CGG TCC CTG TAC ACC AAC CGG TCC AGG CTC ACC CTT GGC TCC AGT GTC ACC TTT G

Col 59R (99 nt): CTT CCT GGG GCT CCG GGC TCT CCA GAG TTA CCC TTC GGT CCA GGT GGG CCA CCA GGT CCT TGA GGT CCG CTT GGG CCT CTA GCT CCT GGA AAG CCT GGG

Col 60R (99 nt): GCT CCT GCG ATC CCT GGT GCA CCGT TGG CGC CTT TGG CTC CTG GTT GAC CGT CGG CAC CTG GGT TTC CAG CGG GAC CAG CGG CAC CAG CAG GAC CCG GC

Col 61R (100 nt): GGC CCA GGC TCA CCA CGC ACT CCT TGA GGT CCT TCG CTA CCC CTT GGC CCC TGC GGA CCT GCC TCA CCC TTA GCT CCA ACG GCA CCA GGG AAT CCA GGA G

Col 62R (100 nt): GAC CGG CGG GAC CAG TAG GAC CAG GTG GAC CGG CCG CTC CTG TAG CAC CAT CGT TTC CCC TGG CAC CAG CTG GAC CGG GCG CTC CTG GTC TAC CTC TCT C

Col 63R (98 nt): ACC GGG CAG GCC TCT TGG TCC CAT CTG ACC TGG CGC GCC GTT TTC TCC TGG AGA TCC TGG CTC ACC CTT GGG TCC TGC TGG TCC AGC GTC CCC CTT AG

Col 64R (99 nt): CTC CGT CTA GAC CGC TGA AAC CCC TGT GTC CCT TCA TGC CTG GCA GTC CAG CTG TCC CCG GCA GAC CCC TAG CGC CCT GCG GTC CTG GTG GTC CCC TCT

Col 65R (99 nt): CTC CTG GCC TCC CGG GCT TAC CAG CCT CGC CGT CGT CTC CAT TCT TAC CAG GTG GTC CCG GAG GTC CCC TTG GGC CCA TAG GTC CAG AGG CTC CCG GCT

Col 66R (100 nt): CTC CAG GTT CTC CAG GTG GTC CTT GAA ATC CCT GAG GTC CAG GAG CAC CAG GTG GGC CAG GCA GTC CCC TTG GAC CGC TCG GAC CCA TGC GGC CAG GCA C

Col 67R (50 nt): GCT GAT ACC TCC GGT GCT CTT CTC GTC GTA GCC GTA GCT CAG CTG AGC GG

5.2 Alignment sequences

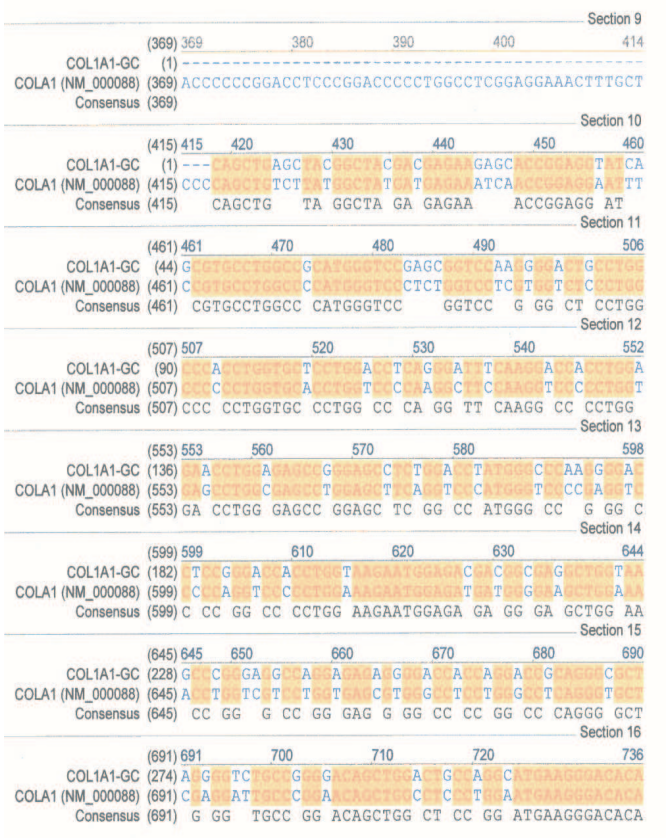

Figure 5.1: Alignment of codon-optimized COL1A1. The codon optimized gene was aligned with the original sequence

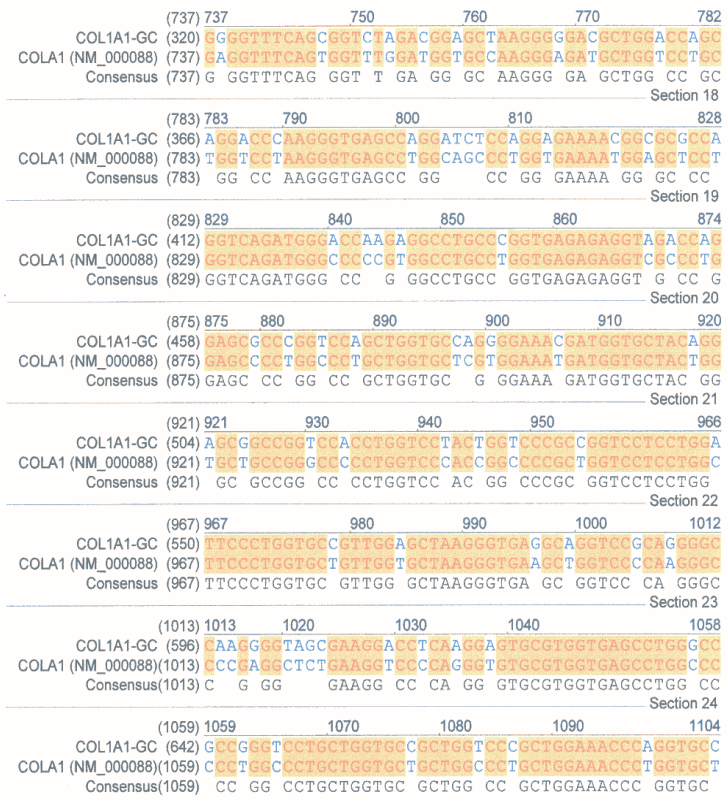

Figure 5.2: Alignment of codon-optimized COL1A1. Part 2

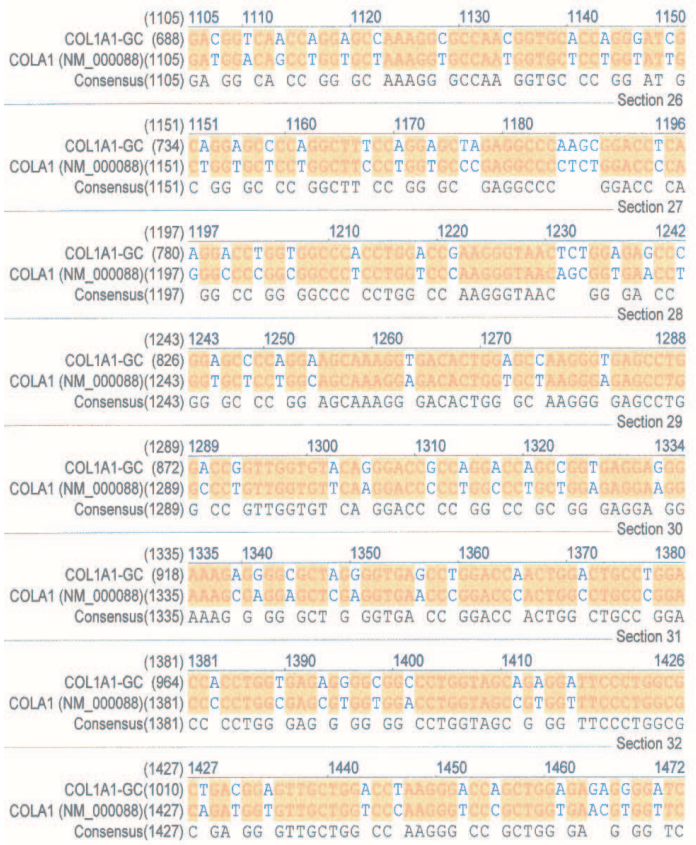

Figure 5.3: Alignment of codon-optimized COL1A1. Part 3

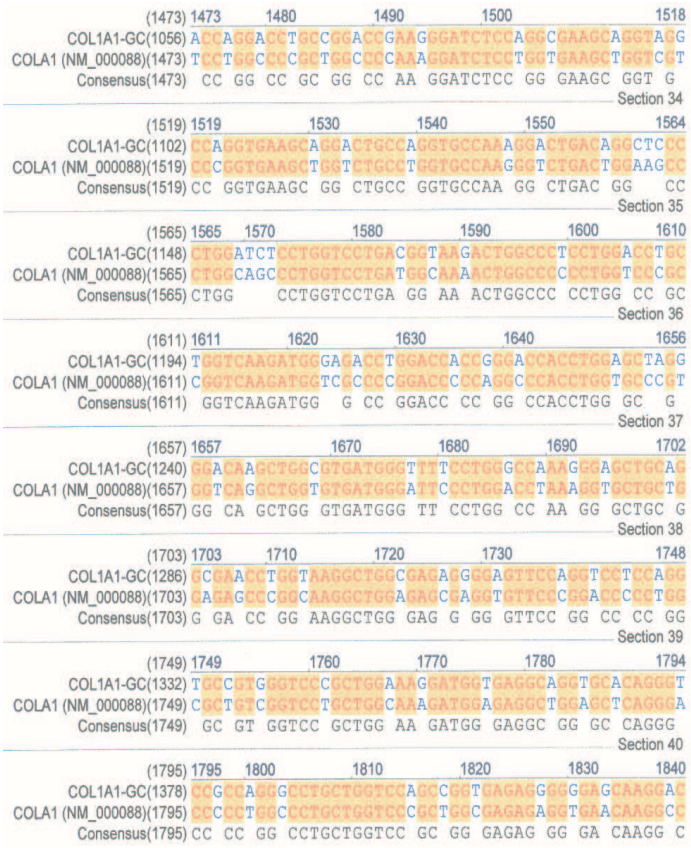

Figure 5.4: Alignment of codon-optimized COL1A1. Part 4

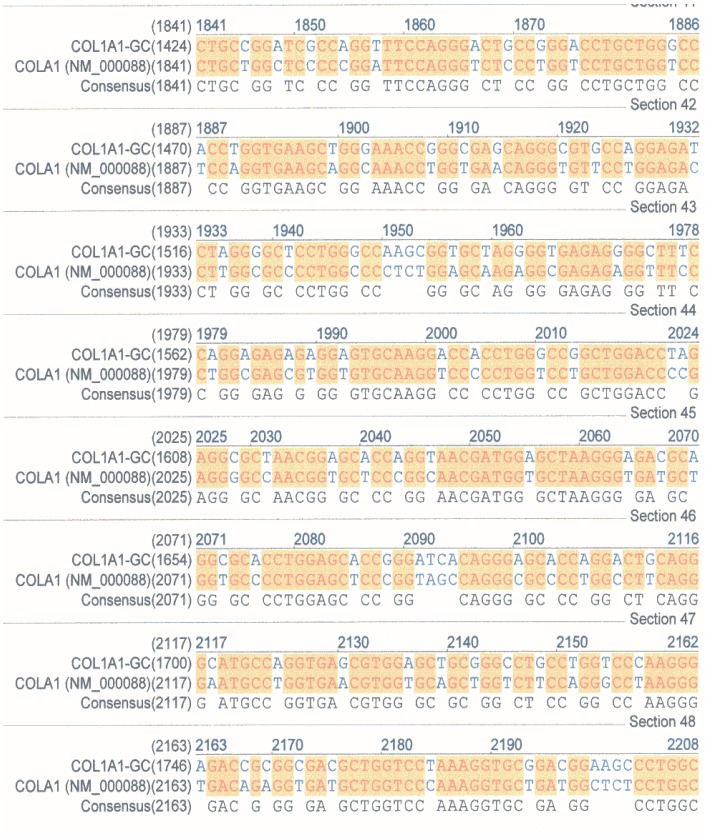

Figure 5.5: Alignment of codon-optimized COL1A1. Part 5

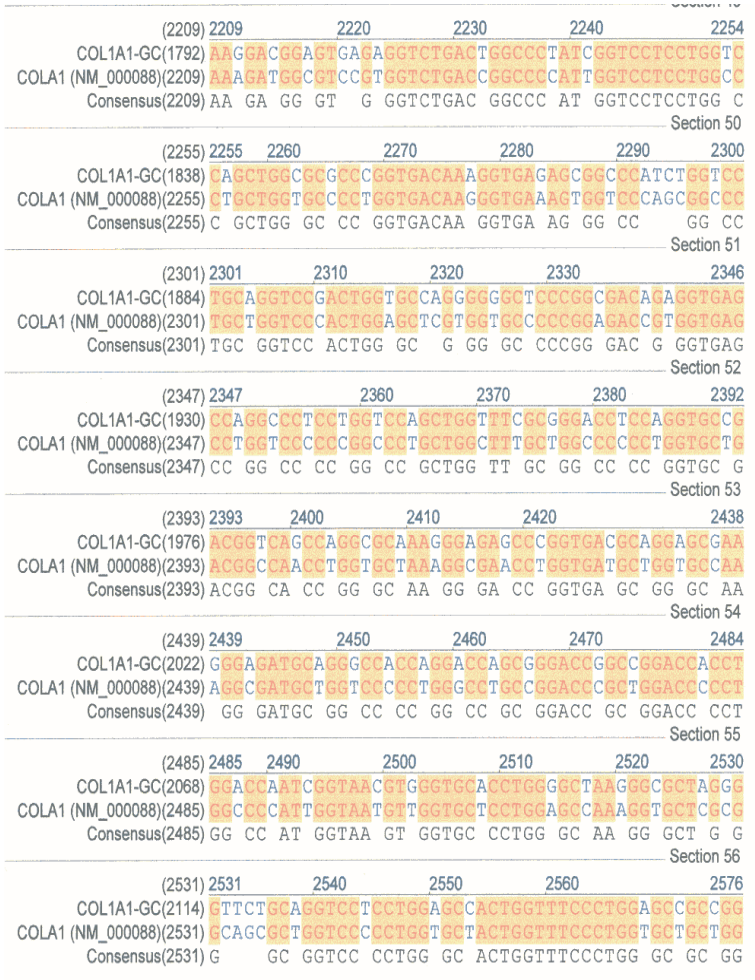

Figure 5.6: Alignment of codon-optimized COL1A1. Part 6

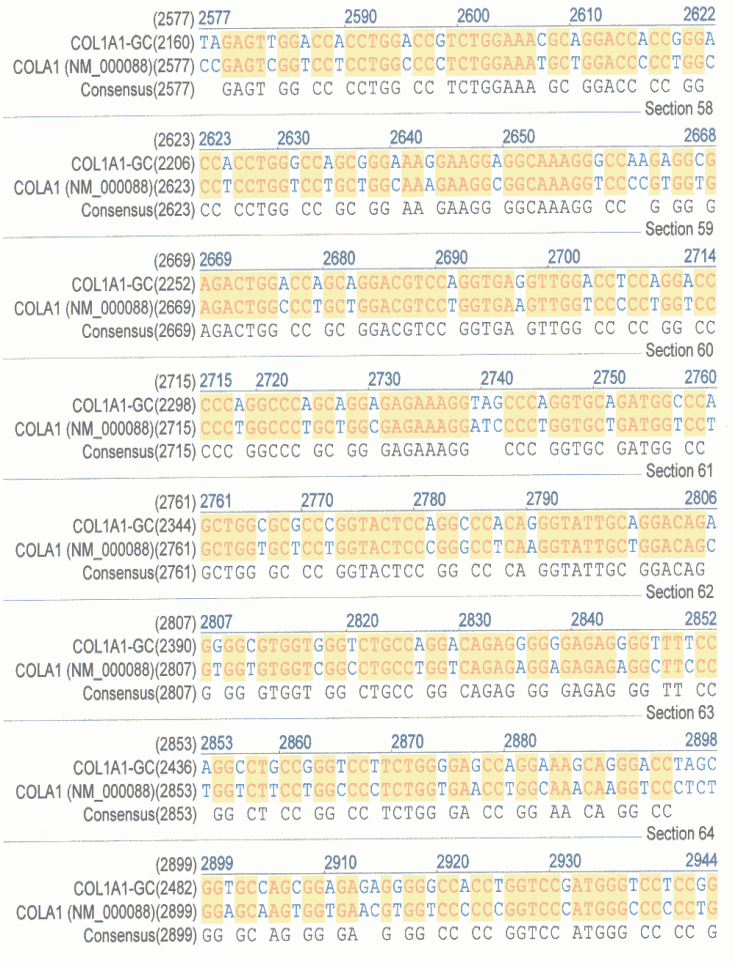

Figure 5.7: Alignment of codon-optimized COL1A1. Part 7

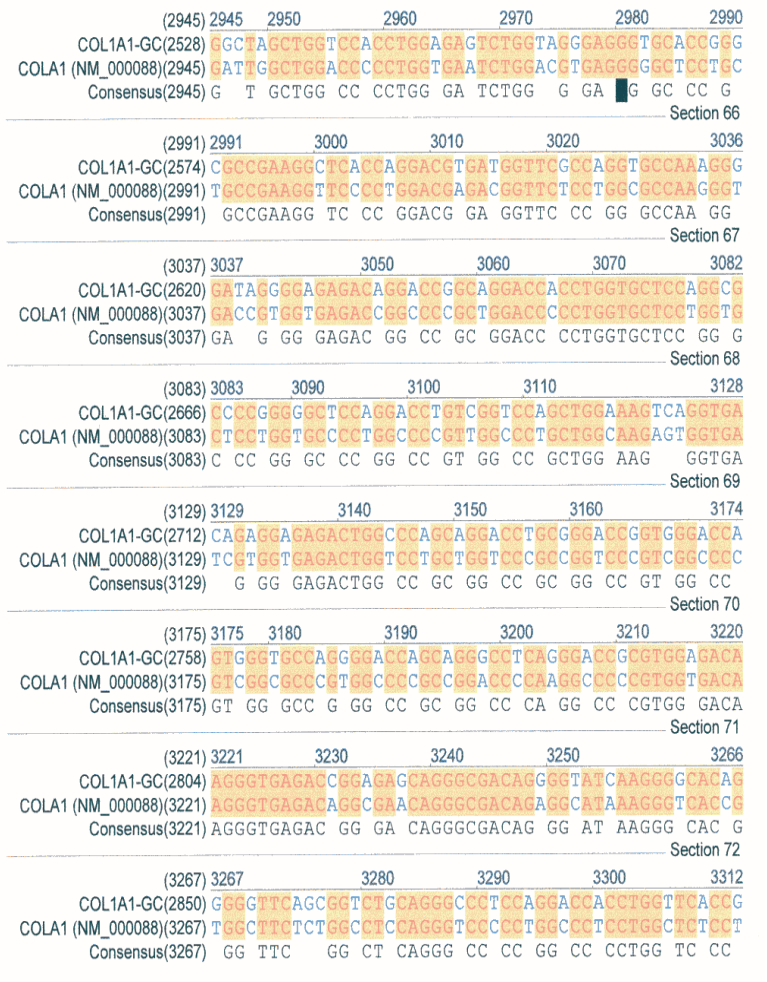

Figure 5.8: Alignment of codon-optimized COL1A1. Part 8

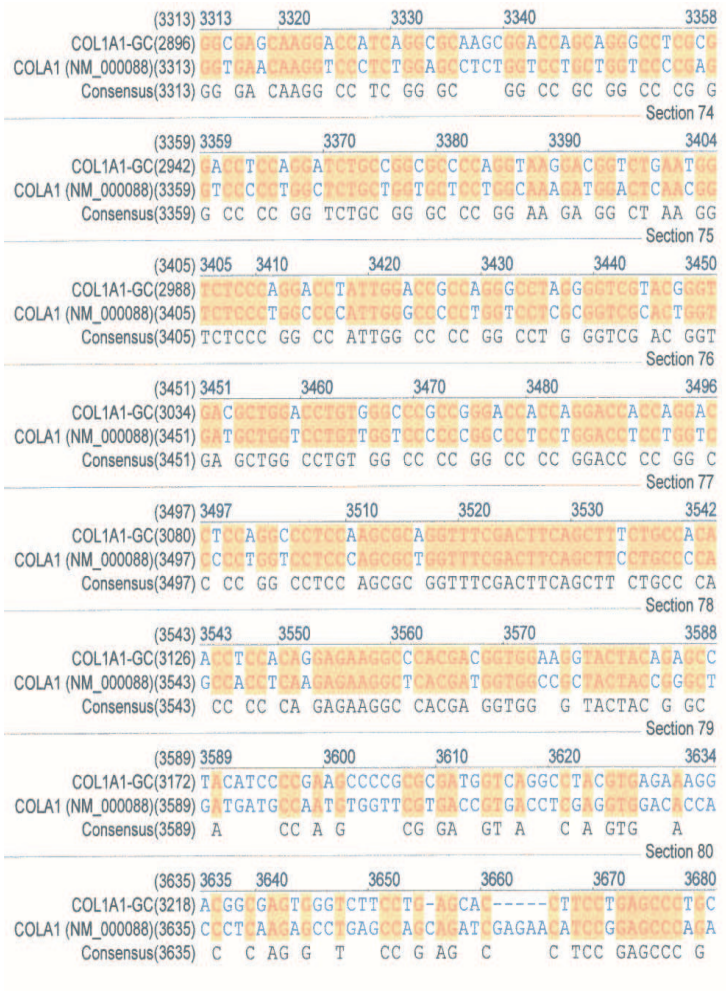

Figure 5.9: Alignment of codon-optimized COL1A1. Part 9

5.3 Primers

5.4 PCR programs

Table 5.1: Primers used for the amplification of intermediate vectors

Primer	Oligos
3'2	CTG GGG CTC CTG CGA TCC CTG GTG C
5'2	CTA AGG GGG ACG CTG GAC CAG CAG G
3'5	CCT TGC CAG GGC TTC CGT CCG CAC C
5'5	GTC CAG CCG GTG AGA GGG GGG AGC AAG
3'8	GCC CCT TGA TAC CCC TGT CGC CCT G
5'8	GGG TCC TTC TGG GGA GCC AGG AAA G
3'10	GCG CGA GCT CTC AGG CAG GGC TCA GG
5'EcoRI	GCG GAA TTC CAG CTG AGC TAC GGC TAC G
3' NotI	CGC CGG CCG CGG CAG GGC TCA GGA AG
5'NcoI/prom	GCG CCA TGG TCC GGA ATC TTC ACC
3'col/sp	GTA GCT CAG CTG AGC GGT GGT GAG AGC
5'sp/col	GCT CTC ACC ACC GCT CAG CTG AGC TAC
3'sp/α	GTA CCA GAT CAT AGC GGT GGT GAG AG
5'sp/α	CTC TCA CCA CCG CTA TGA TCT GGT AC
3'α	GCC GGA CTA GTT CAC TCC AGC TCG C
3'sp/β	CGA GCT GTG AAG CGG TGG TGA GAG C
5'sp/β	GCT CTC ACC ACC GCT TCA CAG CTC G
3'β	CGG ACG CTC CGG AAG AGG AAG ACC

Table 5.2: PCR program used for amplification of collagen fragments

Step	Temperature (°C)	Time (Min)	Number of cycles
Denaturation	97	1	1
Annealing	65	20	
Extension	72	1	
Denaturation	97	0.5	25
Annealing	67	20	
Extension	72	1	
Extension	72	5	

Table 5.3: Direct PCR screening program.

Step	Temperature (°C)	Time (Min)
Denaturation	95	1
Annealing	60	1
Extension	72	1