Purification and characterization of recombinant collagens/gelatins from transgenic corn seeds

by

Cheng Zhang

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Chemical Engineering

Program of Study Committee:
Charles E. Glatz, Major Professor
Lawrence A. Johnson
Patricia A. Murphy
Balaji Narasimhan
Peter J. Reilly

Iowa State University

Ames, Iowa

2008

TABLE OF CONTENTS

ABSTRACT

Corn offers a number of advantages as production host for low- and intermediate-value protein-based products that could be produced in large volume as co-products of corn based biorefining. We selected one such potential product, a recombinant human 100 kDa type I collagen alpha 1 (CIα1) and a 44 kDa single chain fragment (a gelatin). The equivalency to the intended product was established by characterizing the immuno-recognition, molecular weight, amino acid composition and sequence, and post-translational modifications.

Purification for characterization of the recombinant proteins from transgenic corn grain began with protein extraction in 0.1M phosphoric acid buffer. The use of low pH minimized the extraction of host components while allowing efficient recovery of both 44 kDa and 100 kDa CIα1s. The 44 kDa CIα1 was purified by a two-step process consisting of ion-exchange chromatography followed by a gel filtration chromatography, resulting in 70% purity and 60% yield. The N-terminal sequence, amino acid composition and immunoreactivity of the purified 44 kDa CIα1 was similar to that of an analogous 44 kDa CIα1 fragment produced by yeast *Pichia*. The corn-derived 44 kDa CIα1 fragment had an intact protein mass of 44,088 kDa that was within 0.2% of the mass calculated from the intended sequence. 78% coverage of the primary sequence of the corn-derived 44 kDa CIα1 was confirmed with tandem mass spectrometry. Glycoprotein staining of the 44 kDa CIα1 revealed no detectable glycosylation.

The 100 kDa CIα1 was purified to close to 100% purity and 16% yield by using membrane filtration followed by four chromatographic steps (two cation-exchange, one

hydrophobic interaction, and one gel filtration chromatography). The 1.2% hydroxyproline in the corn-derived 100 kDa CIα1, corresponding to about 10% of the hydroxyproline content in native human collagens, confirmed both some ability of corn to hydroxylate proline, but also the necessity of co-expression of a heterologous P4H if higher content is desired. The low hydroxyproline content was well correlated with its relatively low melting temperate (26^oC vs 40^oC in native human collagen). A fraction of the expressed CIα1 formed triple helices resistant to pepsin hydrolysis. A 29 amino acid foldon at the C-terminus was not cleaved from the CIα1 chains but could be removed by pepsin treatment. The amino acid composition and immunoreactivity of the purified CIα1 was as expected and similar to that of an analogous CIα1 produced by the yeast *Pichia*.

The evidence that the CIα1s accumulated in corn grain with the intended composition and post-translational modifications established transgenic corn as a viable way to produce this particular full-length 100 kDa CIα1 and its 44kD CIα1 fragment, and potentially other "designer" gelatins with specific molecular weights and properties tailored to suit various material property specifications.

Two milling processes have been evaluated for their potential to reduce the downstream cost by enriching the CIα1s in fractionated tissues prior to protein extraction and purification. Both dry-milling and wet-milling processes were able to concentrate and capture ca. 60% of the total CIα1s in ca. 20-25% of the total kernel mass. The 44 kDa and 100 kDa CIα1s were enriched 2.5 and 5.8 times respectively in the dry-milled germ-rich fractions compared to whole corn kernel. For the wet milling, the 44 kDa CIα1 was captured in fine fiber-, coarse fiber-, and gluten-rich fractions with 7.7 times enrichment, and the 100 kDa CIα1 was recovered in fine fiber-, coarse fiber-, and germ-rich fractions with 4.6 times

enrichment. In summary, both milling processes can increase the CIα1 purity in selected fractions. While wet and dry milling achieved similar performance in enriching the 100 kDa CIα1, wet milling is superior to dry milling in enriching the 44 kDa CIα1, as the former achieved higher purity in the extract and higher purification factor (PF). CIα1s were associated with both coarse fiber and fine fiber when wet-milled, but the mechanism behind that discovery was not clear. While the main advantage offered by fractionation is the reduction in volume and impurity load to the separation train, an initial on-farm dry-milling step would eliminate the need to transport viable transgenic seeds, thus reducing risk of accidental release.

CHAPTER 1. GENERAL INTRODUCTION

Motivation and Objectives

Corn has been one of the most important crops for food and animal feed and, industrially, for the production of biofuel, adhesives, and sweeteners. Additional uses of corn can come from recombinant technology, which could provide transgenic varieties to produce therapeutics and industrial enzymes at large volume and low cost, potentially as co-products of corn-based biorefining. We selected two such potential high volume products, a recombinant collagen and a single chain fragment of collagen (a gelatin), to demonstrate the feasibility of producing recombinant industrial proteins compatible with economical co-production in a corn-based biorefinery. Over 50,000 metric tons of collagens and gelatins (denatured collagens) are consumed annually for pharmaceutical use, with more than 80% used for capsule manufacturing. Most of the gelatins are derived from bovine or porcine bone or skin as a co-product of the meat production industry. Use of animal-derived products especially in pharmaceutical applications results in allergenicity and safety concerns related to animal-derived contaminants, lack of traceability, and inconsistent product quality due to the variability of tissue origin and animal age. Plants as bio-factories have the potential to eliminate those problems and achieve a safe and consistent product.

The current research was to 1) develop strategies to recover and purify a low expression level recombinant human collagen I alpha 1 (CIα1) chain and a 44 kDa CIα1 fragment from transgenic corn seeds; 2) investigate the accuracy and efficacy of the transgenic corn expression system for the correct production of those two recombinant

proteins; and 3) evaluate the effectiveness of two well-established milling systems (dry and wet milling) for producing kernel fractions enriched in recombinant proteins, thus reducing the downstream process burden and enabling integration of recombinant protein production into a corn biorefinery.

We started with purification and characterization of a 44 kDa CIα1 fragment that was expressed in corn at a level of 20 mg of 44 kDa CIα1/kg corn seed, permitting recovery of sufficient material from the transgenic grain for characterization. We then developed the more challenging recovery of the full-length 100 kDa human CIα1 chain expressed at the very low level of 3 mg 100 kDa CIα1/kg corn seed. We evaluated the use of conventional, non-affinity chromatography to accomplish that task. We focused the characterization of the 100 kDa CIα1 on its functional and structural properties. These results provided information as to how well the corn was capable of processing the synthesized single chains to a mature collagen, especially carrying out the post-translational modification and assembly of the triple-helical structure.

Our industrial partners who produced the transgenic grain had targeted expression of the proteins to the corn germ as a strategy to boost expression level and enable plant tissue fractionation to reduce matrix contaminants in the protein extracts. We evaluated the two well-established milling systems to separate corn kernel into different fractions and monitored the distribution of the two recombinant collagens. The information obtained could be used to assist choosing the best strategy for recovery and purification.

Dissertation Format

This dissertation is arranged in five chapters. References are listed at the end of the chapter in which they are cited. Chapter 1 provides background information relating to collagen fundamentals, recombinant collagen production by yeast and plant expression systems, with emphasis placed on purification process development and protein property characterization. Chapters 2-4 are manuscripts being submitted for publication that describe the experimental work discussed above. Chapter 2 describes the purification and characterization of the 44 kDa CIα1 fragment. Chapter 3 evaluates the expression of the 100 kDa CIα1, with focus placed on the development and optimization of the purification process and evaluation of the expressed protein's structural and functional properties. Chapter 4 compares the distributions of the two recombinant CIα1s in the fractions generated by two well-established milling processes and assesses their suitability for the subsequent extraction and purification process. Finally, the last chapter (Chapter 5) summarizes the overall conclusions and proposes future work.

Literature Review

Native Collagens and Gelatins

Collagens are a family of extracellular proteins with an essential role to maintain the function and structure of the tissues by providing them with mechanical properties. They are amply distributed in bone, cartilage, tendon, cornea, and ligament. There are at least 27 types of collagens identified in humans based on genome information (Baez et al. 2005). Gelatins used commercially are denatured collagens produced from bovine or porcine bones or skins

by acid or base extraction as a co-product of the meat production industry. Over 50,000 metric tons of collagens and gelatins are consumed every year just for medical uses. They are used as biomaterials in pharmaceutical applications, such as capsules for the oral delivery of drugs and vaccine stabilizers, and in tissue engineering applications such as corneal implants (Xie et al. 2001), hemostats and bone grafts (Yang et al. 2004a; Yang et al. 2004b). Other applications are described in several good reviews (Friess 1998; Lee et al. 2001).

Structure

The basic structure of collagen is a triple-helix consisting of three peptide chains, called α chains. Each chain is characterized by repeating Gly-X-Y motifs. X and Y can be any amino acid except glycine, and are often proline or hydroxyproline (Gelse et al. 2003; Myllyharju and Kivirikko 2004). For every three amino acids, a glycine, which is the smallest amino acid, is required to fit inside the triple-helix to achieve compact structure along the central axis of the molecule. For the human type I collagen α1 chain (termed CIα1), which is the target of this work, the triple-helix collagenous domain usually contains about 1000 amino acids and results in 300 nm length. While the triple-helix structures endow the molecules with exceptional tensile strength, the non-collagenous domains located at both ends play important roles in helping form the packed structure (Figure 1). C-propeptide is thought to initiate the formation of triple-helical structure by aligning the C-terminal domains of the three α chains, and the process propagates towards the N terminus. The function of the N-propeptide is not fully understood, and is believed to regulate the size of the fibrils. The non-helical telopeptide is able to assist covalent cross-linking within collagen molecules themselves and between collagen molecules and other surrounding molecules (Gelse et al.

2003).

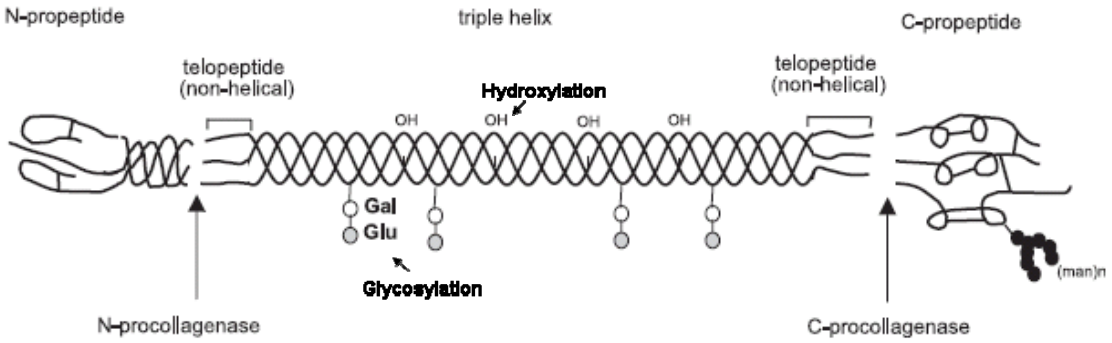

Figure 1. Molecular structure of type I collagen (Reprinted from (Gelse et al. 2003)**,** with permission from Elsevier.)

Collagen Synthesis and Post-translational Modification

Procollagen molecules, which include both triple-helical region, C- and N-telopeptides, and C- and N-propeptides, undergo a series of posttranslational modification after they are directed into the lumen of the endoplasmic reticulum (ER) by the signal peptide which is then usually cleaved off immediately. The conversion to mature collagen (with C- and N-propeptide properly cleaved) relies on the proper post-translational processing of the nascent procollagen. One of the most important modifications is hydroxylation of proline residues which usually occurs on either C3 or C4 sites to form 3-hydroxyproline or 4-hydroxyproline, respectively. While the function of 3-hydroxyproline is unknown, the 4-hydroxyproline is essential to forming the intramolecular hydrogen bonds that provide the thermal stability of the triple-helical structure. In addition, the 4-hydroxyproline assists the formation of intermolecular cross-linking between collagen molecules to develop stable fibrils. The content of the hydroxyproline directly affects the collagen's melting point T_m. About 50% of the proline in bovine type-I collagen are hydroxylated and the corresponding

T_m is ~ 42^{o}C (Merle et al. 2002). The hydroxylation can also take place on lysine residues. The hydroxylysine residues contribute to the fibril integrity and represent the potential site for the attachment of sugar groups (Gelse et al. 2003). The sugar components are either the monosaccharide galactose or the disaccharide of galactose and glucose, linked to the hydroxylysine residues. The function of the hydroxylysine glycosylation is not yet fully understood, but it is thought to affect fibril formation (Myllyharju 2005). It is in the ER that procollagens form the triple-helical structure with the assistance of the C-propeptide.

After procollagen assembly and processing, the triple-helical molecules are secreted into the extracellular space via the Golgi apparatus, and they continue to undergo processing and modification to form fibrils, which is directed by the information encoded in the collagen structures. Both C-propeptides and N-propeptides are cleaved by specific processing enzymes during or following secretion, the N and C-proteinases (Koide and Nagata 2005), and the cleavage of the C-propeptide is thought to be essential for regulating the fibril self-assembly. Baez's work showed that expression without N-propeptides gave a correct folding, suggesting the N-propeptide did not play an important role in collagen assembly (Baez et al. 2005). Fibril structures are further stabilized by the formation of cross-linking between collagen molecules, and it is the hydroxylysine residues in the telopeptides that are responsible for the cross-linking formation (Gelse et al. 2003). The correct formation of fibrils is indispensable for collagens to maintain their functions such as high tensile strength.

Recombinant Collagen and Gelatin Production

Use of animal-derived products, especially in pharmaceutical applications, results in allergenicity and safety concerns related to animal-derived contaminants, lack of traceability,

and inconsistent product quality due to the variability of tissue origin and animal age. Those products represent potential sources pathogens (Asher 1999) and allergens. An allergic reaction was observed on injection of the α2 chain of bovine type I collagen (Sakaguchi et al. 1999). In addition, the animal-derived collagens, especially the gelatins, are mixtures of proteins lacking uniformity in size, pI (isoelectric point), structure, composition and gelling points. They also show lot-to-lot variability. The safety concerns and unpredictability associated with the animal-derived collagens and gelatins prompt the development of transgenic plant sources without those drawbacks.

Pichia-derived Recombinant Collagens and Gelatins

While the accumulation of recombinant collagen and of collagen fragments has been demonstrated in many other recombinant expression systems such as mammalian cells (Fertala et al. 1994; Fichard et al. 1997; Geddis and Prockop 1993), insect cell culture (Myllyharju et al. 1997; Nokelainen et al. 1996), *Escherichia coli* (Cappello 1990; Goldberg et al. 1989), *Bacillus* (Kajino et al. 2000), silkworms (Tomita et al. 2003), milk of transgenic animals (John et al. 1999; Toman et al. 1999) and other yeasts (*Sacchromyces cerevisia*e (Olsen et al. 2001; Vaughan et al. 1998), *Hansenula polymorpha* (de Bruin et al. 2002; Geerlings et al. 2007), the current preferred recombinant technology for collagen and gelatin production is based on *Pichia* yeast fermentation (Baez et al. 2005; Olsen et al. 2003). FibroGen Inc., a biopharmaceutical company based in South San Francisco CA, has demonstrated commercial production of *Pichia*-derived recombinant collagens and gelatins for pharmaceutical application. Since type I collagen is the most abundant type of collagen, the following discussion will mostly focus on the type I collagen.

During the synthesis of recombinant collagens, the majority of procollagens expressed in *Pichia pastoris* cannot be secreted into the extracellular medium, but are found to accumulate within the ER of the cells even when the authentic signal sequence of the collagen was replaced with the *S. cerevisiae* α mating factor pre-pro sequence; The latter was reported to be a more efficient signal sequence than the authentic one (Keizer-Gunnink et al. 2000; Pakkanen et al. 2006; Vuorela et al. 1997). Both the triple-helical conformation and large size of the procollagens (trimer, ~270 kDa) were found to adversely affect secretion (Pakkanen et al. 2006). It is necessary to lyse the cells by external forces to release the procollagens, which are then converted to collagens by the controlled treatment with proteases, commonly human pepsin to remove C-and N-propeptides. By careful manipulation of gene types and optimizing fermentation processing parameters, expression levels higher than 1 g/l can be achieved (Baez et al. 2005).The recombinant collagens produced from *Pichia pastoris* contain a majority of monomers, with only 5-10% cross-linked dimers and a minor amount of trimers. Compared to the animal-derived collagens that usually consist of more than 50% dimers, trimers and higher order oligomers, the recombinant collagens produced from *Pichia pastoris* show much higher homogeneity (Olsen et al. 2003).

Prolyl 4-hydroxylase (P4H), a $\alpha_2\beta_2$ tetramer commonly found in vertebrates, is believed to play a crucial role in forming stable triple-helical collagens, as well as fibril structure. P4H can hydroxylate proline residues, and as mentioned before, the resulting hydroxyproline is capable of forming intramolecular hydrogen bonding within the collagen molecules and assisting the development of intermolecular cross-linking within fibrils (between collagen molecules), key steps for a stable conformation. However, no P4H activity could be found in either yeasts or bacteria, which means the recombinant collagens expressed

could be in a non-triple-helical and non-functional state. Co-expression of P4H gene with collagen, therefore, is usually required in yeasts to get the correct triple-helical structure (Vuorela et al. 1997). The successful co-expression of human P4H with recombinant human type I-III collagens has been demonstrated in the yeast *P. pastoris* where two subunits of P4H formed an active $\alpha_2\beta_2$ tetramer, which effectively hydroxylated the procollagen chains resulting in the correct triple-helical structure. The mature collagen showed hydroxyproline content and T_m almost identical to those of the corresponding non-recombinant counterparts, and self-assembled into native-type fibrils at predefined conditions (Myllyharju et al. 2000; Nokelainen et al. 2001). Although C-propeptides (~ 250 amino acids) are considered to be essential for correct triple helix structure assembly, substitution of C-propeptide with a short bacterial peptide (a 29 amino acid sequence) not only led to more effective assembly of triple- helical structure in human type I and III collagen molecules, but also resulted in higher expression level (Pakkanen et al. 2003). The significance of the finding suggests the expression system could be engineered to develop correctly processed recombinant collagens for high-level production.

By cloning or expressing part of the collagen gene, recombinant gelatins with well-defined molecular weights and pIs could be produced to match specific requirements or applications. Accumulation levels up to 14 g/L of secreted gelatin have been reported in *Pichia* fermentations (Werten et al. 1999). These collagen-related proteins have been produced with consistent quality and used successfully for many applications. Studies using a series of *Pichia*-derived recombinant human type I collagen α1 chain fragments ranging in size from 56 to 1014 amino acids indicated that these recombinant collagen fragments can replace animal-derived gelatin on many current medical applications (Olsen et al. 2003). In

addition, a low molecular weight 8.5 kDa *Pichia* derived gelatin (pI 9.4) was shown to be equivalent to animal-derived gelatin as a stabilizer of an influenza virus vaccine with good lot-to-lot reproducibility and low allergenic potential (Baez et al. 2005). One challenge to producing recombinant gelatins via *Pichia* fermentation is the high susceptibility of the gelatins to proteolysis due to their open unfolded structure. The proteolytic degradation could be minimized but not eliminated by running the fermentation at pH 3 (Baez et al. 2005; Werten et al. 1999). A custom-designed, highly hydrophilic recombinant gelatin, however, has been shown to be fully intact by SDS-PAGE, N-terminal sequencing, gel filtration chromatography and mass spectrometry (Werten et al. 2001), suggesting recombinant gelatins could be produced as intended. Since the structure is not a requirement in recombinant gelatins, no co-expression of P4H is needed and the generated gelatins were not hydroxylated (Werten et al. 2001). Unlike the collagens, the recombinant gelatins can be secreted into the extracellular medium by introduction of a heterologous signal sequence such as the *S. cerevisiae* α-mating factor pre-pro sequence (Baez et al. 2005; Pakkanen et al. 2006; Werten et al. 1999; Werten et al. 2001). Secretion eliminates the need of cell lysis by physical disruption and reduces the downstream process burden.

Transgenic Plant-derived Recombinant Collagens and Gelatins

Transgenic tobacco has been investigated as a cost-effective host to produce recombinant collagens. Although the plant has an endogenous P4H, the enzyme has different substrate specificity than the mammalian counterpart and the plant is thus unable to correctly hydroxylate recombinant human collagens. The human type I collagen accumulating in tobacco without cloned animal P4H shows that the content of hydroxylated proline is at a

negligible level (Ruggiero et al. 2000). The non-hydroxylated collagen, however, is able to correctly fold and form a triple-helical structure but at much slower propagation rate, compared with hydroxylated bovine type I collagen. Due to the lack of hydroxylation, the tobacco-derived collagen exhibits increased flexibility and reduced thermal stability. The latter is reflected in its reduced T_m, 30.5^{o}C, which represents a more than 10^{o}C decrease in temperature compared to 42^{o}C for the native bovine Type I collagen. Moreover, the lack of hydroxylation prevents self-assembly into fibrils in the physiological buffer conditions where the corresponding non-recombinant bovine collagens do self-assemble. The phenomenon further emphasizes the unquestionably critical role of proline hydroxylation in stabilizing the collagen structure (Perret et al. 2001; Ruggiero et al. 2000).

As in *P. pastoris*, co-expression of human collagen with heterologous P4H is necessary in the transgenic tobacco if hydroxylated recombinant collagens are intended (Merle et al. 2002). Recombinant hydroxylated human type I collagen has been successfully produced from transgenic tobacco by the co-expression of a cloned animal P4H enzyme. The tobacco-derived collagen has a hydroxyproline content of 8.4%, a little bit lower than the 10% in *Pichia pastoris*- derived collagens and native collagens. In addition, the collagen adopts the expected triple-helical conformation, and has a T_m of 37^{o}C, which is several degrees lower than that of the native bovine collagen, and much higher than the 30.5^{o}C for the tobacco-derived non-hydroxylated collagens. The differences in the hydroxyproline content may explain the disparity in the T_m (Merle et al. 2002).

Both non-hydroxylated and hydroxylated recombinant collagens undergo spontaneous processing from procollagens to mature collagens by the endogenous tobacco proteases during extraction and purification, thus eliminating the conversion of procollagen

by using pepsin digestion, as employed in the production of *Pichia pastoris*-derived collagens (Merle et al. 2002; Ruggiero et al. 2000). The result demonstrates a potential advantage of plants for the large-scale, low-cost production of collagens. Besides transgenic tobacco, recent studies indicated that transient and stable expression of human type I collagen can be achieved in barley and rice suspension cell culture (Dodge et al. 2006; Holkeri et al. 2005) and in barley seeds (Ritala et al. 2007).

Transgenic Plants for Recombinant Protein Production

Since the first report of successful transformation of plants in 1983 (Fraley et al. 1983), significant advances have been made in plant biotechnology to move plant-derived recombinant proteins from laboratory research to market. To date, three research-grade recombinant proteins have already been on the market: avidin, β-glucuronidase, and trypsin, all developed by ProdiGene Inc. from transgenic corn (Horn et al. 2004b). Although no recombinant plant-derived pharmaceutical protein for human use has yet reached the market, more than ten plant-derived proteins, such as gastric lipase and lactoferrin produced in maize by Meristem Therapeutics (Clermont Ferrand, France), were at one time in clinical trials (Ma et al. 2005; Sparrow et al. 2007). In January 2006 the first registration for a tobacco-derived poultry vaccine was granted to Dow AgroSciences (Indianapolis, IN) by the USDA, with the expectation of the first recombinant product for human use entering the market by 2009 (Sparrow et al. 2007). Several recent reviews describe the merits of using plant hosts and possible expression strategies for the production of bioindustrial or biopharmaceutical recombinant proteins (Boehm 2007; Ma et al. 2003; Menkhaus et al. 2004a; Sparrow et al. 2007; Twyman et al. 2003).

Transgenic plants offer several advantages over other recombinant systems. They are considered to be much safer than both microbe and animal expression systems because they generally lack human pathogens, oncogenic DNA sequences, and endotoxins. Another major advantage of transgenic plants is the potential for lower capital investment and production cost, and the latter was thought to be less than 15% of that required for mammalian cell culture (Menkhaus et al. 2004a). In addition, recombinant plant systems are easy to scale up: production volumes can be increased by increasing planted acreage (Howard and Hood 2005). If recombinant proteins are targeted for expression in seeds, the transgenic plants are particularly attractive in that they can accumulate the proteins in a small volume and provide a stable storage environment. Seed may be stored without loss of protein activity for at least six months (Bai and Nikolov 2001; Stoger et al. 2000) or more than a year in some cases (Kusnadi et al. 1998). The existence of well-established infrastructure for cultivating, harvesting, transporting, and storing plant matter also contributes to low production cost. In general, the low cost of growing a large amount of biomass, ease of scale-up (increase of planted acreage), capability of long term storage, and established practices for efficient processing of the plant make transgenic plant systems an ideal option for industrial recombinant protein production. However, regulatory issues have offset these technical and economic advantages, stalling implementation of plant-made proteins. The risk of contamining the feed and food chain with transgenic food crops, such as rice, barley, maize and soybean, is currently not acceptable whether or not the recombinant protein poses any actual health risk. Prevention of contamination must be addressed by temporal and spatial segregation measures to control cross-pollination with other crops during flowering cycles

and elimination of seed spillage and seed mixing during transportation and handling in order to allay those regulatory concerns (Sparrow et al. 2007).

Protein Localization in Plant Cells

Previously we mentioned that one major advantage of transgenic plants as bioreactors is the low capital and running costs. Since the expression level directly affects the cost of cultivation, extraction, purification and waste disposal, and consequently, the feasibility of commercialization, the economic advantages of plant production systems cannot be realized unless the recombinant proteins can be produced at high yields. It is estimated that the expression level needs to be at least 0.1% of biomass weight (e.g. 1 g recombinant protein/ 1 kg of seeds) for plants to remain competitive in cost (Nandi et al. 2005). One of the most important factors governing the yields of recombinant proteins is the protein localization in plant cells because the subcellular compartments where recombinant proteins accumulate strongly influence the interrelated processes of folding, assembly and post-translational modification. All of those contribute to protein stability and thus determine the final yields (Fischer et al. 2004).

A tissue-specific promoter (e.g. embryo-specific, endosperm-specific) can be chosen that allows a particular tissue, such as germ or endosperm of corn seeds, to accumulate a particular recombinant protein. In addition to the promoter, a signal sequence fused to the expression construct can target the protein to a subcellular compartment within that tissue to maximize that promoter's work. If a signal sequence is present, the newly synthesized proteins are transported to the ER. From there, the proteins are directed to the subcellular compartments such as Golgi apparatus, cell wall, vacuole, and plasma membrane. Proteins

can also be retained in the ER, like the recombinant human type I collagen expressed in the transgenic tobacco (Merle et al. 2002; Ruggiero et al. 2000). In the absence of a signal sequence, the recombinant protein will accumulate in the cytoplasm (Hood 2004; Streatfield 2007).

In a study to achieve high-level expression of a recombinant fungal laccase in transgenic maize, combinations of three distinct promoters (constitutive, embryo-specific, and endosperm-specific) and four targeting sites (cell walls, ER, vacuole, and cytoplasm) were tested to determine best locations for high-level accumulation of the protein. The embryo (germ)-specific expression with the protein being targeted to the cell wall resulted in highest expression level (Hood et al. 2003). The same technology was applied to the expression of a sweet protein brazzein, and achieved a high expression level-4% of the total soluble protein in maize seed (Lamphear et al. 2005).

Other experiments have shown that recombinant antibodies targeted to the subcellular compartments through the ER or retained in the ER usually accumulate to levels that are several orders-of-magnitude greater than those of antibodies expressed in the cytosol (Schillberg et al. 2003; Twyman et al. 2003). The reason is that many important post-translational processing elements like chaperones, protein-disulfide isomerase, and glycosylation enzymes are present in the ER but not at all or scarcely present in the cytosol. Those elements are required for correct protein folding and assembly. Without them, the proteins do not assemble into the mature structure and consequently degrade, resulting in low final yields (Boehm 2007). Protein storage vacuoles, a post-Golgi compartment, are also reported to be good destinations as stable protein accumulation is facilitated (Stoger et al. 2005). A recombinant enzyme (1, 3–1, 4)-β-glucanase) accumulates at a high level in

transgenic barley storage vacuoles of the mature grain endosperm (Horvath et al. 2000). All those results have demonstrated that optimizing transformation and cell lines to target to different subcellular localization might consistently elevate the expression levels to those of microbial or animal-based systems.

Targeting of recombinant proteins to subcellular compartments, such as oil bodies or the plasma membrane, not only increases expression level as described previously, but facilitates isolation and purification as well. One example is the oleosin-fusion platform developed by SemBioSys Genetics (http://www.sembiosys.com, Calgary, Canada), in which the target recombinant hirudin was fused with oleosin. The oleosin-hirudin fusion protein was correctly targeted to the oil body membrane and separated from the majority of other seed proteins by flotation centrifugation. The recombinant hirudin was released from its fusion partner by endoprotease digestion (Parmenter et al. 1995). Another example is an anti-viral antibody fragment fusion protein which accumulated at the plasma membrane and could be extracted in a small volume by proper buffers and detergents (Schillberg et al. 2000).

Similar to targeting recombinant proteins to subcellular compartments, using tissue-specific promoters to target proteins to a specific part or tissue that can be easily separated from the rest of the plant material before protein extraction would not only result in higher purity extract, but also facilitate the downstream process and reduce the processing cost. For example, maize germ (embryo) is considered well suited for the expression of some recombinant proteins at high levels, as demonstrated by the expression of a recombinant β-glucuronidase (Witcher et al. 1998) , avidin (Hood et al. 1997), fungal laccase (Hood et al. 2003), bovine trypsin (Woodard et al. 2003), brazzein (Lamphear et al. 2005), and bovine aprotinin (Zhong et al. 2007). Since the germ makes up about 12% of the whole maize kernel,

a germ-rich fraction generated by the dry-milling process can result in a 5-10 fold enrichment of the recombinant protein content, as well as a 5-10 fold reduction of biomass compared to the whole grain. As the cost of processing is inversely proportional to the concentration of the product protein in the starting materials and the reduced biomass leads to the decrease in extract volume, the fractionation of seeds by dry milling can greatly reduce the cost. In addition, the 90% of non-germ fraction (basically endosperm and hulls) can be utilized for other applications such as ethanol or adhesive production, and domestic animal feed. Therefore, the cost of the recombinant protein could be further reduced by the byproduct credits (Howard and Hood 2005).

Post-translational Modification

The protein synthesis pathway is conserved between plants and animals so plants appear to efficiently fold and assemble recombinant human proteins. This is a great advantage since proteins failing to fold properly will be degraded, resulting in low yields. The ability of plants to fold and assemble complex proteins correctly is demonstrated by their capacity to produce functional serum antibodies and secretory antibodies (Twyman et al. 2003). However, there are some differences in posttranslational modifications. As discussed before, transgenic tobacco is unable to correctly modify/hydroxylate recombinant human collagens unless a heterologous P4H is co-expressed (Merle et al. 2002). Another difference is glycosylation. First, plant glycosylation patterns and extents differ from those of animals (Giddings et al. 2000). Plants might add plant-unique sugar groups to or alter the sugar moieties on the recombinant proteins during the post-translational processing. A maize-derived bovine trypsin was unexpectedly glycosylated, while the native bovine trypsin did

not have glycosylation. But the difference did not prevent the protein from functioning equivalently to its native counterpart (Woodard et al. 2003). The β1,2-xylose and α1,3-fucose of plant origin were found to replace the galactose and sialic acid in the maize-derived human lactoferrin (Samyn-Petit et al. 2001), and a similar case was reported for rice-derived human lactoferrin (Nandi et al. 2005). Although the unintended addition and alteration of carbohydrate groups in plants represents potential allergen or antigen sources, no cases of allergenic or immune responses have been reported to date (Horn et al. 2004b; Howard and Hood 2005). Another glycosylation difference between plant and mammalian systems is that the former cannot produce sialic acid for inclusion in the glycoprotein (Howard 2005). Addition of the sialic acid is thought to be responsible for the extended clearance time of the glycoproteins in the blood stream. Consequently the difference might limit the potential of the transgenic plants to produce some recombinant human proteins (Howard 2005).

Corn as Host for Recombinant Protein Production:

Corn or maize (*Zea Mays* L) is the most important crop in the United States and one of the most important crops in the world. Corn is widely distributed throughout the world and accepted as a safe product as food, feed, and raw material for industrial applications. The protein content of the corn varies depending on type, variety, and other factors, ranging from 6-18% in the kernel. The proteins can be divided into four categories according to the Osborne classification, which is based on the solubility of the proteins in various solvents. The Osborne classification scheme is as follows (adapted from (Johnson 2000; Lawton and Wilson 2003).

Albumins: soluble in pure water (salt-free)

Globulins: soluble in dilute salt solutions, but insoluble in pure water

Prolamins: soluble in non-polar solvents, e.g. 70% ethanol

Glutelins: soluble in dilute acid or base w/o a small amount of reducing agent

The corn kernel can be separated either into chemical constituents (starch, gluten/protein, fiber, germ/oil) or anatomical parts (endosperm, germ and bran), depending on different objectives. The endosperm accounts for more than 80% of the whole corn kernel on dry weight basis, and contains the majority of the starch. Although endosperm has an average protein content of only 8-9%, it contains about 75% of total proteins in the corn due to its large volume. The germ makes up about 11% of the whole corn kernel, contains most of the oil, and has the greatest protein concentration (17 to 20%). The bran accounts for 5% of the whole corn kernel and is mainly composed of fiber (Johnson 2000). A detailed distribution in each fraction is shown in Table 1.

Corn (or maize) has been the most widely used plant host for the production of recombinant proteins. In addition to the advantages possessed by all plant production systems, such as low cost of growing biomass, the convenience of scale-up, well-established practices and infrastructure for efficient harvesting, transporting, storing and processing, it provides extra benefits of stable long-term storage encased seed with little or no microbial burden, well-developed transformation technology and ease of *in vitro* manipulation (Ma et al. 2003; Ramessar et al. 2008; Sparrow et al. 2007; Twyman et al. 2003). Those are among the major factors that prompted Prodigene Inc. to choose corn as the first plant for the commercial production of three recombinant proteins-avidin, β-glucuronidase, and trypsin (Horn et al. 2004b). In 2005, Meristem Therapeutics was granted authorization to use corn for field production of mammalian gastric lipase for the treatment of lipid malabsorption related to

Table 1. A typical composition in the anatomical components (adapted from (Johnson 2000).

Component	Protein, %	Fat, %	Starch, %	Fiber, %	Other, %
Endosperm	8.0	0.8	87.6	3.2	0.4
Germ	18.4	33.2	8.0	14.0	26.4
Bran	3.7	1.0	7.3	83.6	4.4
Whole kernel	8.7	4.1	71.3	3.0	1.5

exocrine pancreatic insufficiency and monoclonal antibodies for the treatment of various types of cancer (www.meristem-therapeutics.com). Other corn-derived recombinant proteins developed within the past several years include aprotinin (Azzoni et al. 2002; Zhong et al. 1999), laccase (Bailey et al. 2004; Hood et al. 2003), brazzein (Lamphear et al. 2005), human lactoferrin (Samyn-Petit et al. 2001), human proinsulin (Farinas et al. 2005b), cellulase (Hood et al. 2007), antibodies (Hood et al. 2002) and vaccines (Horn et al. 2004a; Lamphear et al. 2004; Lamphear et al. 2002; Streatfield et al. 2003). The main disadvantages of corn are its open pollination and wide planting for food and feed. Appropriate confinement measures are necessary to prevent the risk of cross-pollination or seed mixing with other corn crops. Planting location and timing can avoid cross-pollination with other corn planting (Ramessar et al. 2008). Until such practices become accepted, lemsa, rice, and barley will continue to be examined as alternatives that have less risk of cross-pollination.

Downstream Processing of Recombinant Proteins Produced in Plants

Downstream processing of recombinant proteins produced in plants usually consists of plant tissue fractionation, protein extraction and protein purification (Menkhaus et al.

2004a). It is estimated that the cost of downstream processing accounts for more than 80% of production cost (Evangelista et al. 1998). Therefore, development and optimization of downstream processing methods is necessary to reduce cost, which, in combination with expression level, might govern the economic feasibility of transgenic plant systems.

Plant Tissue Fractionation

Since transgenic corn seeds were used in our research, the following discussion mainly focuses on corn seeds. The main purpose of fractionation is to reduce the biomass for the protein extraction. Dry milling is able to separate the corn kernel into three anatomical fractions: germ, endosperm, and hulls. In dry milling the corn kernels are moisture-conditioned, dehulled, degermed, roller milled and fractionated through a series of sieves and aspirators (Yildirim et al. 2002). Fractionation by dry milling has a significant impact on downstream processing. For example, if the desired protein is targeted to germ, the dry-milling process can generate a germ-rich fraction with a 5- to 10- fold enrichment of the protein content compared to the whole grain (Howard and Hood 2005). In addition, the biomass required to produce the desired protein is reduced 5-to 10- fold which not only greatly lowers the extraction cost, but also cuts down the cost of waste disposal. But the high oil content of the germ-rich fraction could interfere with protein purification; therefore, defatting /deoiling may be necessary before protein extraction and purification. In addition, germ represents a more complex matrix for protein accumulation than endosperm, which could make the separation of target proteins from host corn proteins more difficult. On the other hand, if the target is the endosperm-rich fraction, a neutral salt buffer would leave nearly 80% of native corn proteins in the residual endosperm solids rather than in the extract,

thus considerably simplifying the later purification process (Menkhaus et al. 2004a). Wet milling, which can separate the corn kernel into a great number of components more cleanly, has not been studied extensively for the downstream processing of recombinant proteins.

Protein Extraction

Protein extraction is carried out by adding ground corn (either germ-rich or endosperm-rich fraction) to aqueous buffers, where the proteins are released from the plant cells to the aqueous medium. The choices of the extraction medium and conditions, such as buffers, ionic strength, pH, ratio of the extraction buffer to the plant tissues, extraction time etc., are important to achieve satisfactory recovery of the recombinant proteins while minimizing the amount of the native proteins extracted, and thus reducing the cost of the later purification processing.

Of all the extraction parameters, pH and ionic strength have the greatest effects on the protein extraction efficiency, and have been most extensively studied. The total extractable corn endogenous protein concentration increased about 10-fold by increasing the pH from 3 to 10 (Azzoni et al. 2002; Zhong et al. 2007). A similar trend in pH was reported in extracting a recombinant human proinsulin from transgenic maize endosperm (Farinas et al. 2005b). Ionic strength has less significant effect than pH, but the total extractable corn endogenous protein concentration also increases with addition of NaCl up to 300 mM (Azzoni et al. 2002). The effects of pH and ionic strength on the recombinant protein extracted from the corn seeds vary, and the optimal extraction conditions should be determined on case-by-case basis. However, as a rule of thumb, to reduce the protein complexity of the extract and simplify the purification process, acidic pH condition is desired,

as long as the recombinant proteins can be solubilized under acidic conditions. In addition, the acidic pH results in less extraction of other corn metabolites, such as carbohydrates and phenolic compounds, which are considered to interfere with chromatography (Farinas et al. 2005a). Since gelatins and collagens are acid soluble (Ruggiero et al. 2000), the above rule can apply here, and acidic pHs will help reduce the crude extract complexity.

The buffers used to maintain the pH during the extraction process necessarily change with the choice of pH. Commonly used extraction buffers include glycine, sodium acetate, sodium phosphate, potassium phosphate, Tris and sodium carbonate with concentrations of 50 mM or below; hence, the contribution to ionic strength is low. There is no report that the buffer salt influences extraction separately from the pH and ionic strength contributions. The ratio of the ground corn seeds to buffers is usually 1:5, or sometimes 1:10, and extraction time needs to be determined empirically, depending on the specific recombinant proteins. Some recombinant proteins may be susceptible to proteolytic degradation due to the concomitant release of the endogenous proteases during the protein extraction. The addition of protease inhibitors or lowering pH could mitigate the degradation (Werten et al. 1999).

Protein Purification:

As with the purification of recombinant proteins produced from other expression systems, no generic procedure is available for purifying a recombinant protein produced from transgenic corn or other plants. A body of knowledge must be built to purify a particular recombinant protein, and it is different for each protein. Therefore, a purification process needs to be designed according to the specific target protein, however, there are still some general rules that can be followed. Purification can generally be divided into three steps

(Figure 2): a capture step, where proteins are partially enriched and purified; all intermediate step, where a majority of contaminant proteins are removed, resulting in a purity of 60-80%; and a polishing step, where trace amounts of contaminant proteins are removed, leading to purities up to 100%. However, there are no definite lines between the steps, and purification can be achieved with two steps or even one step. The number of steps used will always depend upon the purity requirements and intended use for the protein.

Chromatography

Chromatography is ubiquitous in protein separations and considered as a very important part of almost any purification strategy due to its high resolution. Chromatography commonly used for each step and the mechanism for each chromatographic method are summarized in Table 2.

(Note to Reader: The following detailed discussion of each chromatographic method is based on my personal understanding and lab experience and is intended for future students. Other readers may want to proceed to the next section on the "Overall Process".)

Affinity chromatography (AC) makes use of highly specific molecular recognition. It has the highest specificity and selectivity of all chromatographic methods, and is the most efficient method especially for target proteins at low concentration in the starting materials. For immunoglobulin (e.g. IgGs), commercially available affinity resins, such as Protein A or Protein G, are ready and easy for use. While for non-immunoglobulin recombinant proteins, a preparation of antibodies for the target proteins (viewed as antigen) is necessary. The best way to obtain the antibodies is by immunizing animals (e.g. mouse, rabbit, and goat) by injecting the product protein purified by other techniques, such as chromatography.

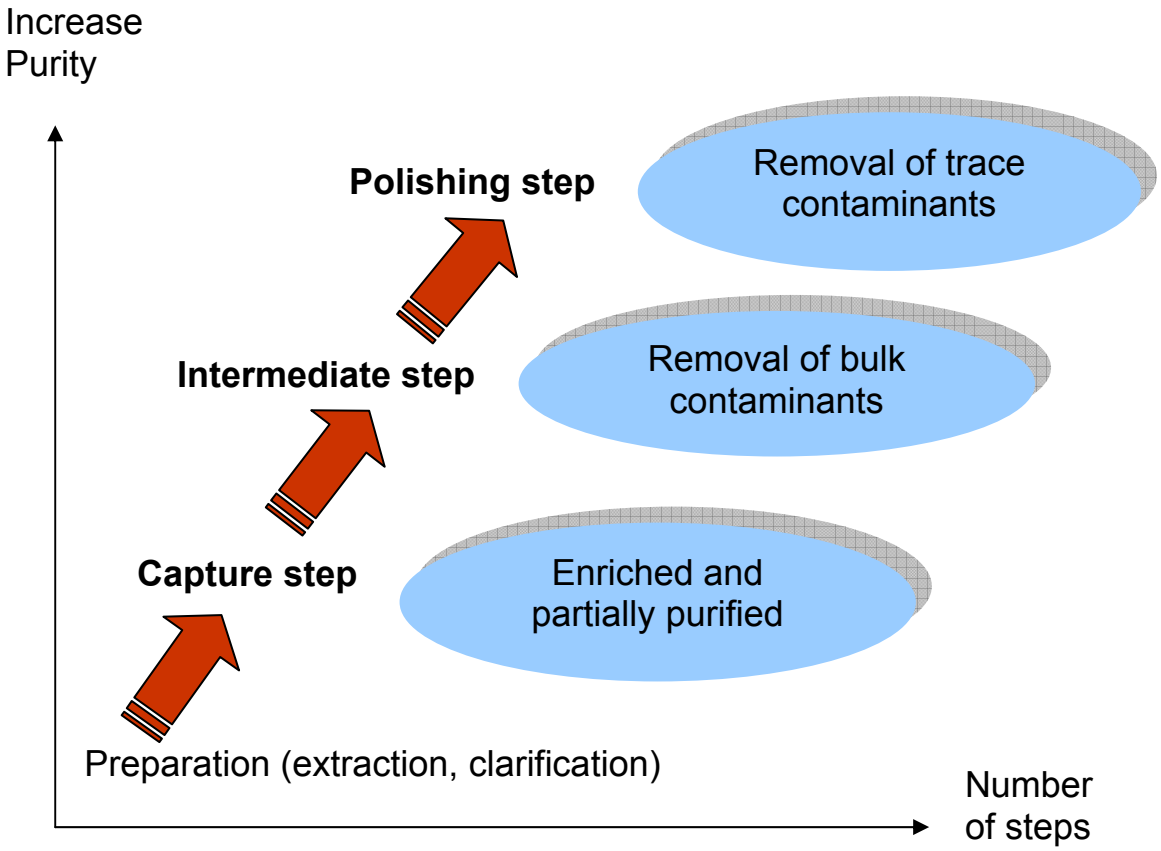

Figure 2. Three step purification strategy

Table 2. Separation technologies in the design of a downstream process (adapted from Protein Purification Handbook (18-1132-29), Amersham Biosciences)

Chromatography	Protein property	Capture	Intermediate	Polishing
AC (Affinity Chromatography)	Biospecificity	√ √ √	√ √ √	√
IEX (Ion Exchange)	Charge	√ √ √	√ √ √	√ √ √
HIC (Hydrophobic Interaction Chromatography)	Hydrophobicity	√ √	√ √ √	√
SEC (Size Exclusion Chromatography)	Size		√	√ √ √
RPC (Reverse Phase Chromatography	Hydrophobicity		√	√ √ √

(Note: more √ means more likely that chromatography can be used in that step)

Commercially available antibodies generated from native proteins could be a convenient option, but it runs the risk that the proteins produced from recombinant systems might not recognize them. The antibodies then need to be immobilized (covalently attached) to a porous matrix, followed by packing into a chromatographic column. One commonly used matrix is highly cross-linked agarose beads which usually have a size <100μm. Detailed information of immobilization methods could be found in an affinity chromatography handbook (Amersham Biosciences, 18-1022-29). Antibody binding to target proteins during chromatography is carried out at neutral pHs, and elution usually performed by using acidic pHs (< 4). Since the acidic pH would denature the proteins, eluted pools (containing target proteins) should be quickly neutralized with buffers like 2M Tris base (pH>9). Elution can also be conducted by increasing salt concentrations or introducing chaotropic agents (e.g. urea), but these eluents are less common than pH elution. A single step of affinity chromatography of the clarified corn extract resulted in a more than 90% pure corn-derived avidin with > 90% recovery (Kusnadi et al. 1998). In addition, affinity chromatography as a capture step separated a corn-derived aprotinin from the majority of host contaminating proteins, increasing its purity from 0.28 to 9.6% with a purification factor of 33 (Azzoni et al. 2002). Without an affinity column, more chromatography steps would be needed to achieve the same purity.

Ion-exchange (IEX) chromatography separates proteins primarily according to their overall charges. The principle of IEX is based on the competition between charged proteins and salt ions for the adsorbent sites. During sample loading, proteins displace salt ions and attach to the resin; at elution, the proteins are displaced by ions at much higher concentration, hence “ion exchange”. The net surface charge of proteins varies according to the surrounding

pH. When above its isoelectric point (pI at which a protein has a net charge of zero) a protein will bind to a cation-exchange (CEX) resin (e.g. carboxymethyl, sulphopropyl resin); when below its pI a protein will behind to an anion exchange (AEX) resin (e.g. diethylaminoethyl). So, the pI of the protein will determine what pH and resins can be used for the purification. For example, a recombinant gelatin with pI of 9.4 will be positively charged at pHs lower than its pI, which makes CEX chromatography a good choice (Olsen et al. 2005).

IEX buffer is necessary to maintain the pH at which the chromatography is running. Correct choice of buffer can be crucial to success since the ionic interactions are highly dependent on pH and buffer species may take part in the ion exchange process. When choosing a buffer, several things need to be considered (Scopes 1994): 1) the pK_a (the pH at which the ratio of conjugate base and acid forms is 1 and the buffer has maximum ability to resist changes in pH) of the buffer should be not more than 1 unit, and preferably 0.5 unit away from the working pH; 2) the buffer species should not interact with the resin, e.g. for AEX resin, choosing a positive buffering ion (Tris) and for a CEX resin, choosing a negative buffering ion (acetate, phosphate) ;3) the pH in the microenvironment of the resin matrix may be significantly different from that of applied buffer due to ion exchanging, so enough buffer species (at least > 20 mM) is required. Another important practical aspect is the actual column/resin capacity for the proteins being adsorbed. The adsorptive or binding capacity of IEX resins varies for individual proteins, highly dependent on properties such as protein size and chromatographic running conditions such as flow rate. So, a preliminary experiment to estimate dynamic binding capacity is usually needed. For optimal separations when performing gradient elution, it is advised to use approximately one-fifth to one-twentieth of the total binding capacity of the column (Scopes 1994). In general, IEX is the most

commonly used chromatography; it has high loading capacity-independent of volume and high resolving power. It concentrates target proteins which are generally stable in running conditions. It allows for versatile operation; optimal separation could be achieved by manipulating pH, buffers, and elution gradients.

Both hydrophobic interaction chromatography (HIC) and reverse phase chromatography (RPC) separate proteins based on the reversible hydrophobic interaction between a protein and the surface of a chromatographic resin. The difference between the two chromatography is that RPC uses organic solvent (e.g. acetonitrile) to elute proteins, and it has high resolution and is more proper to separate small proteins or peptides for analytical purposes, while HIC uses aqueous solution for elution, and its goal is to purify proteins of interest for further analysis or investigation. Since the downstream process development is to prepare pure materials, the following discussion will focus on HIC.

The hydrophobicities of target proteins influence the selection of hydrophobic resin which is usually made of linear aliphatic chains or aromatic groups. The longer the chain, the more hydrophobic the resin is.

Protein samples in most cases need to be in a high salt concentration that enhances the interaction in order for the target protein to bind, however, too high a salt concentration would start precipitating the protein. Therefore for the HIC development, an important goal is to find a buffer condition such that the ionic strength/salt concentration is high enough to promote adsorption, but not so high as to cause precipitation of the product protein. The most commonly used buffer at laboratory scale is ammonium sulfate, but sodium sulfate and sodium chloride are more common in industry because they cause less severe disposal problems. Usually simple batch adsorption tests are sufficient to determine the amount of

salts needed. Proteins are often eluted by decreasing the salt concentration. If the target protein binds strongly and does not elute even at low salt concentration, then it is worth considering a less hydrophobic resin. An alternative is adding organic solvent (e.g. ethylene glycol gradient up to 50%), chaotropic species (urea, guanidine hydrochloride) or detergents to reduce eluent polarity (Protein Purification Handbook (18-1132-29), Amersham Biosciences). But this will increase the complexity of the eluted sample, and removal of those solutes can be a nuisance.

Since HIC loads sample with high salt and separates proteins by the mechanism complementary to IEX, a natural and convenient two-step chromatographic process could be to start with an IEX column followed by HIC (Lightfoot and Moscariello 2004). The IEX column can directly bind proteins from the feed stream, and the higher salt needed for elution is helpful for adsorption on the HIC column. In general, HIC does not usually have as sharp separation as IEX, but it has similarly high loading capacity. It does not need buffer exchanging or pH adjustment, but adding sufficient salt to ensure binding before sample application is necessary.

Size exclusion chromatography (SEC) separates proteins according to their size, which is closely related to their MWs. When applied to the separation of biomolecules in aqueous solutions, it is referred to as gel filtration chromatography (GFC). The development of SEC process is straightforward. No buffer adjustment of sample from previous steps is needed for sample application. Buffer conditions are chosen to meet the requirements for further purification, analysis or storage, as buffer composition does not directly affect resolution. Up to 0.15 M NaCl can be added to aqueous buffers to avoid non-specific ionic

interaction of proteins with the column matrix. Proteins are collected in purified form in the chosen buffer (Gel filtration: principles and methods (18-1022-18), Amersham Biosciences).

However, there are some weaknesss of SEC that limit its use at large scale. Applied sample volume is very low, up to 4% of total column volume to achieve good resolution, which in turn requires a relatively large amount of chromatographic resin. The follow rate is low, usually < 0.5 cm/min, leading to prolonged process time for a large column. In general, SEC is ideal for the final polishing step on bench scale when sample volumes are greatly reduced. It is a very gentle method, and often gives near 100% recovery (Scopes 1994). In addition, the SEC process is easy to develop and operate.

Overall Process

Purification of recombinant proteins from transgenic plants is usually composed of a combination of the above discussed chromatographic methods. Two chromatographic steps (affinity followed by CEX chromatography) produced a corn-derived aprotinin with high purity (single SDS-PAGE band) and about 40% yield (Azzoni et al. 2005). Purification of the corn-derived aprotinin was also tried by a non-affinity process, which started with CEX followed by HIC. The results demonstrated that the combination of CEX and HIC was a promising alternative to affinity chromatography for purifying recombinant aprotinin from the germ fraction of transgenic corn seeds (Zhong et al. 2007). More generic processes composed of multi-step chromatographic columns were used to purify several other proteins. A recombinant human lysozyme was purified to a single SDS-PAGE band with 60% recovery from a high-expression transgenic rice line by a combination of CEX and SEC (Huang et al. 2002). In some cases, more than three chromatographic steps were required to

achieve satisfying purity, especially when the target protein purity was very low in the starting material. A four- step chromatography procedure (AEX I, HIC, AEX II, SEC) was reported to yield 90% plus pure β-glucuronidase from corn extract with 27% recovery (Kusnadi et al. 1998). Coming with more chromatographic steps are decreasing yield and increasing cost, which further emphasizes the importance of extraction optimization and pretreatment (e.g. plant tissue fractionation) to reduce the complexity of the crude sample solution.

Non-chromatographic Methods

Although chromatography has been the workhorse for purifying high-value recombinant pharmaceutical proteins, such as monoclonal antibodies, the high cost and limited throughput associated with chromatography prompted the development of new or alternative technologies for large-scale recombinant protein production (Przybycien et al. 2004; Thommes and Etzel 2007). The traditional, inexpensive, non-adsorptive method of precipitation is one promising alternative, especially for the production of low-valued proteins as byproducts from less-expensive raw materials, such as plants. In addition, precipitation can be complementary to adsorptive chromatography in terms of separation mechanism.

Many precipitation methods including heat, organic solvent, salt, isoelectric, and polyelectrolyte precipitation have been evaluated as early purification steps. Heating the corn extract was reported to remove a fraction of native corn proteins from corn extract (Zhong et al. 2007), but would not be useful for heat-labile products. Organic precipitants, such as acetone and ethanol, have been used to recover recombinant gelatin produced from yeast *P.*

pastoris (Werten et al. 1999; Werten et al. 2001) and *H. polymorpha* (de Bruin et al. 2002). Like heating, addition of organic solvents has a tendency to denature proteins, which makes it less commonly used. Salts are commonly used to precipitate proteins, and ammonium sulfate and sodium chloride have been successfully applied to recombinant gelatin or collagen expressed in yeasts and plants (Baez et al. 2005; Nokelainen et al. 2001; Perret et al. 2001; Werten et al. 2001). Isoelectric precipitation from canola resulted in a 3-fold enrichment of lysozyme (Zaman et al. 1999), and from pea showed a 4-fold enrichment of β-glucuronidase with 80% yield (Menkhaus et al. 2004b). Polyelectrolyte precipitation with PEI, a polycationic precipitating agent, showed different level of success: GUS precipitated from pea achieved 18-fold enrichment, whereas GUS from corn only resulted in a purification factor of 2.6 (Menkhaus et al. 2002). In addition, polyelectrolyte precipitation of lysozyme from canola gave a poor performance. A disadvantage of using polyelectrolyte is that it is not easily removed from precipitated complex. In general, the efficacy of precipitation methods varies greatly, and depends highly on the composition of the protein mixture.

Cost-effective Process Development for Recombinant Collagen Purification

Compared to other high-valued recombinant proteins such as antibodies, recombinant gelatins and collagens are usually used as supplementary ingredients. So for the sole purpose of their purification, relatively low purity is acceptable (>80%) and cost-effective purification methods are critical to make them competitive with their native counterparts. Literature has shown that salt precipitation alone is able to achieve high purity for plant or *Pichia*-derived gelatins or collagens. A tobacco-derived recombinant collagen with an estimated

concentration of 0.02 mg/ml in tobacco extract was purified to 90% by 0.9 M sodium chloride (NaCl) precipitation (Ruggiero et al. 2000). Another example was the use of differential ammonium sulfate precipitation to obtain a gelatin with an estimated purity of > 98% (Werten et al. 2001). In addition, Baez et al. (2005) reported that a precipitation protocol including three salt precipitations at different pHs resulted in highly purified *Pichia*-derived collagens. These results suggest that gelatins or collagens are sensitive to salt precipitation, and this information might help us to establish a similar precipitation method for the cost-effective production of corn-derived gelatins or collagens.

Protein Characterization

Characterization of recombinant proteins from plant hosts is needed to assess accuracy of the expression system. SDS-PAGE still remains a useful tool to check the purity and MW of the product protein. With the help of the specific immuno-reaction between antibody and antigen, Western blotting, a tool complementary to SDS-PAGE, is able to confirm the existence of the target protein and detect the isoforms or fragments of the protein. While SDS-PAGE can only estimate a protein's mass, MS (mass spectrometry) is able to measure mass with high accuracy, which provides supplemental evidence that the target protein is correctly expressed. Since the inaccuracy of MS is proportional to the size of the molecule, proteins with relatively high MW (e.g. > 50 kDa) might not allow the MS to pick up the mass difference between the measured protein mass and theoretical mass and attribute that difference to any specific modification. In order to better identify the protein, tandem MS (MS-MS) can be implemented to deduce internal peptide sequence. The extent to which the MS data represents the protein sequence is referred to as "coverage". Literature use of

MS to verify that a protein is being expressed relies on as little as 10-15% sequence coverage on the basis of the low probability of another protein containing the same combination of peptides (Liebler 2002). Higher sequence coverage is needed if we would like to detect protein modification on amino acids. N-terminal sequencing would provide more evidence to identify the protein and common N-terminal modifications.

Previous studies of plant-derived recombinant proteins have focused more on the production and downstream processing. Limited protein characterization such as SDS-PAGE and N-terminal sequencing was provided for aprotinin (Azzoni et al. 2002), β-glucuronidase, and avidin (Kusnadi et al. 1998) for assessing the accuracy of expression and product stability. More characterization, including Western blotting, MS, and glycosylation staining, was implemented for a corn-derived bovine trypsin (Woodard et al. 2003) to better analyze the product. The differences in molecular weight between corn seed-derived and native trypsin were determined to be due to O-linked glycosylation on the corn-derived protein demonstrated by the chemical glycosylation removal resulting in the proteins having the same apparent molecular weight. Glycoprotein staining can be used to detect the presence of glycosylation. However, detailed information about the glycan groups, and at what amino acids modification takes place requires combination of advanced analytical techniques such as HPLC and MS, and the discussion of that is beyond the scope of this review. Karnoup et al. (2005) showed that O-linked glycosylation found in a corn-derived human IgA took place through hydroxylated proline residues, indicating that plant was able to convert proline to hydroxyproline. Amino acid analysis is a convenient method to determine how much proline is hydroxylated, as long as the protein is pure and a hydroxyproline standard is available.

Purification of Recombinant Gelatins or Collagens for Characterization

For the purpose of characterizing recombinant gelatins or collagens from *Pichia* or tobacco, a general approach to prepare a pure sample is a combination of precipitation method plus a single step of chromatography such as ion exchange (Olsen et al. 2001; Perret et al. 2001) or gel filtration chromatography (Myllyharju et al. 2000; Nokelainen et al. 2001). Some characterization such as amino acid analysis needs ultrapure preparation, so one more step of high resolution chromatography such as reverse-phase chromatography (Baez et al. 2005; Perret et al. 2001; Ruggiero et al. 2000) or affinity chromatography (Vuorela et al. 1997) is commonly used. A possible alternative to separate minor amounts of impurities is SDS-PAGE, which is nicely compatible with analytic methods such as MS and N-terminal sequencing.

References

Asher DM. 1999. The transmissible spongiform encephalopathy agents: concerns and response of the United States regulatory agencies on the origin. Dev. Biol. Stand. 100 103-118.

Azzoni AR, Kusnadi AR, Miranda EA, Nikolov ZL. 2002. Recombinant aprotinin produced transgenic corn seed: Extraction and purification studies. Biotechnol. Bioeng. 80(3):268-276.

Azzoni AR, Takahashi K, Woodard SL, Miranda EA, Nikolov ZL. 2005. Purification of recombinant aprotinin produced in transgenic corn seed: Separation from CTI utilizing ion-exchange chromatography. Braz. J. Chem. Eng. 22(3):323-330.

Baez J, Olsen D, Polarek JW. 2005. Recombinant microbial systems for the production of human collagen and gelatin. Appl. Microbiol. Biotechnol. 69(3):245-252.

Bai Y, Nikolov ZL. 2001. Effect of processing on the recovery of recombinant beta-glucuronidase (rGUS) from transgenic canola. Biotechnol. Prog. 17(1):168-174.

Bailey MR, Woodard SL, Callaway E, Beifuss K, Magallanes-Lundback M, Lane JR, Horn ME, Mallubhotla H, Delaney DD, Ward M and others. 2004. Improved recovery of active recombinant laccase from maize seed. Appl. Microbiol. Biotechnol. 63(4):390-397.

Boehm R. 2007. Bioproduction of therapeutic proteins in the 21st century and the role of plants and plant cells as production platforms. Ann. N.Y. Acad. Sci. 1102: 121–134.

Cappello J. 1990. The biological production of protein polymers and their Use. Trends Biotechnol. 8(11):309-311.

de Bruin EC, Werten MW, Laane C, de Wolf FA. 2002. Endogenous prolyl 4-hydroxylation in *Hansenula polymorpha* and its use for the production of hydroxylated recombinant gelatin. FEMS Yeast Res. 1(4):291-298.

Dodge C, Baez J, Sudarshana M, McDonald K. 2006. Production of recombinant human gelatin in rice cell cultures. AAPS Annual Meeting, San Antonio, TX. W5024.

Evangelista RL, Kusnadi AR, Howard JA, Nikolov ZL. 1998. Process and economic evaluation of the extraction and purification of recombinant beta-glucuronidase from transgenic corn. Biotechnol. Prog. 14(4):607-614.

Farinas CS, Leite A, Miranda EA. 2005a. Aqueous extraction of maize endosperm: Insights for recombinant protein hosts based on downstream processing. Process Biochem. 40(10):3327-3336.

Farinas CS, Leite A, Miranda EA. 2005b. Aqueous extraction of recombinant human proinsulin from transgenic maize endosperm. Biotechnol. Prog. 21(5):1466-1471.

Fertala A, Sieron AL, Ganguly A, Li SW, Alakokko L, Anumula KR, Prockop DJ. 1994. Synthesis of recombinant human procollagen-Ii in a stably transfected tumor-cell line (Ht1080). Biochem. J. 298:31-37.

Fichard A, Tillet E, Delacoux F, Ruggiero F. 1997. Human recombinant alpha 1(V) collagen chain - Homotrimeric assembly and subsequent processing. J. Biol. Chem. 272(48):30083-30087.

Fischer R, Stoger E, Schillberg S, Christou P, Twyman RM. 2004. Plant-based production of biopharmaceuticals. Curr. Opin. Plant Biol. 7(2):152-158.

Fraley RT, Rogers SG, Horsch RB, Sanders PR, Flick JS, Adams SP, Bittner ML, Brand LA, Fink CL, Fry JS and others. 1983. Expression of Bacterial Genes in Plant-Cells. Proc. Nat. Acad. Sci .USA-Biol. Sci. 80(15):4803-4807.

Friess W. 1998. Collagen - Biomaterial for drug delivery. Eur. J. Pharm. Biopharm. 45(2):113-136.

Geddis AE, Prockop DJ. 1993. Expression of human Col1a1 gene in stably transfected Ht-1080 cells - the production of a thermostable homotrimer of type-1 collagen in a recombinant system. Matrix 13(5):399-405.

Geerlings TH, de Boer AL, Lunenborg MGJ, Veenhuis M, van der Klei IJ. 2007. A novel platform for the production of nonhydroxylated gelatins based on the methylotrophic yeast *Hansenula polymorpha*. FEMS Yeast Res. 7(7):1188-1196.

Gelse K, Poschl E, Aigner T. 2003. Collagens - structure, function, and biosynthesis. Adv. Drug Delivery. Rev. 55(12):1531-1546.

Giddings G, Allison G, Brooks D, Carter A. 2000. Transgenic plants as factories for biopharmaceuticals. Nat. Biotechnol. 18(11):1151-1155.

Goldberg I, Salerno AJ, Patterson T, Williams JI. 1989. Cloning and expression of a collagen-analog-encoding synthetic gene in Escherichia-coli. Gene 80(2):305-314.

Holkeri H, Wahlstrom E, Ritala A, Baez J, Makinen K, Nuutila AM. 2005. Production of non-hydroxylated recombinant collagen in barley. J. Biotechnol. 118: S84-S84 (Abstract).

Hood EE. 2004. Where, oh where has my protein gone? Trends Biotechnol. 22(2):53-55.

Hood EE, Bailey MR, Beifuss K, Magallanes-Lundback M, Horn ME, Callaway E, Drees C, Delaney DE, Clough R, Howard JA. 2003. Criteria for high-level expression of a fungal laccase gene in transgenic maize. Plant Biotechnol. J. 1(2):129-140.

Hood EE, Love R, Lane J, Bray J, Clough R, Pappu K, Drees C, Hood KR, Yoon S, Ahmad A and others. 2007. Subcellular targeting is a key condition for high-level accumulation of cellulase protein in transgenic maize seed. Plant Biotechnol. J. 5(6):709-719.

Hood EE, Witcher DR, Maddock S, Meyer T, Baszczynski C, Bailey M, Flynn P, Register J, Marshall L, Bond D and others. 1997. Commercial production of avidin from transgenic maize: characterization of transformant, production, processing, extraction and purification. Mol. Breed. 3(4):291-306.

Hood EE, Woodard SL, Horn ME. 2002. Monoclonal antibody manufacturing in transgenic plants - myths and realities. Curr. Opin. Biotechnol. 13(6):630-635.

Horn ME, Pappu KM, Bailey MR, Clough RC, Barker M, Jilka JM, Howard JA, Streatfield SJ. 2004a. Advantageous features of plant-based systems for the development of HIV vaccines. J. Drug Target. 11(8-10):539-545.

Horn ME, Woodard SL, Howard JA. 2004b. Plant molecular farming: systems and products. Plant Cell Rep. 22(10):711-720.

Horvath H, Huang JT, Wong O, Kohl E, Okita T, Kannangara CG, von Wettstein D. 2000. The production of recombinant proteins in transgenic barley grains. Proc. Natl. Acad. Sci. USA 97(4):1914-1919.

Howard JA. 2005. Commercialization of biopharmaceutical and bioindustrial proteins from plants. Crop Sci. 45(2):468-472.

Howard JA, Hood E. 2005. Bioindustrial and biopharmaceutical products produced in plants. Adv. Agron. 85: 91-124.

Huang JM, Nandi S, Wu LY, Yalda D, Bartley G, Rodriguez R, Lonnerdal B, Huang N. 2002. Expression of natural antimicrobial human lysozyme in rice grains. Mol. Breed. 10(1-2):83-94.

John DCA, Watson R, Kind AJ, Scott AR, Kadler KE, Bulleid NJ. 1999. Expression of an engineered form of recombinant procollagen in mouse milk. Nat. Biotechnol. 17(4):385-389.

Johnson LA. 2000. Corn: The major cereal of the Americas. In: Kulp K, Ponte JG, editors. Handbook of Cereal Science and Technology. 2nd ed. New York. p 31-80.

Kajino T, Takahashi H, Hirai M, Yamada Y. 2000. Efficient production of artificially designed gelatins with a *Bacillus brevis* system. Appl. Environ. Microbiol. 66(1):304-309.

Karnoup AS, Turkelson V, Anderson WHK. 2005. O-Linked glycosylation in maize-expressed human IgA1. Glycobiology 15(10):965-981.

Keizer-Gunnink I, Vuorela A, Myllyharju J, Pihlajaniemi T, Kivirikko KI, Veenhuis M. 2000. Accumulation of properly folded human type III procollagen molecules in specific intracellular membranous compartments in the yeast *Pichia pastoris*. Matrix Biol. 19(1):29-36.

Koide T, Nagata K. 2005. Collagen biosynthesis. Top. Curr.Chem. 247:85-114.

Kusnadi AR, Hood EE, Witcher DR, Howard JA, Nikolov ZL. 1998. Production and purification of two recombinant proteins from transgenic corn. Biotechnol. Prog. 14(1):149-155.

Lamphear BJ, Barker DK, Brooks CA, Delaney DE, Lane JR, Beifuss K, Love R, Thompson K, Mayor J, Clough R and others. 2005. Expression of the sweet protein brazzein in maize for production of a new commercial sweetener. Plant Biotechnol. J. 3(1):103-114.

Lamphear BJ, Jilka JM, Kesl L, Welter M, Howard JA, Streatfield SJ. 2004. A corn-based delivery system for animal vaccines: an oral transmissible gastroenteritis virus vaccine boosts lactogenic immunity in swine. Vaccine 22(19):2420-2424.

Lamphear BJ, Streatfield SJ, Jilka JA, Brooks CA, Barker DK, Turner DD, Delaney DE, Garcia M, Wiggins B, Woodard SL and others. 2002. Delivery of subunit vaccines in maize seed. J. Control. Release 85(1-3):169-180.

Lawton JW, Wilson CM. 2003. Chapter 9, Proteins of the Kernel in Corn: Chemistry and Technology White PJ, Johnson LA, editors. Minneapolis, MN. AACC Press.

Lee CH, Singla A, Lee Y. 2001. Biomedical applications of collagen. Internat. J. Pharmaceut. 221(1-2):1-22.

Liebler DC. 2002. Introduction to proteomics: tools for the new biology. Totowa, NJ Humana Press. p167-184.

Lightfoot EN, Moscariello JS. 2004. Bioseparations. Biotechnol. Bioeng. 87(3):259-273.

Ma JKC, Barros E, Bock R, Christou P, Dale PJ, Dix PJ, Fischer R, Irwin J, Mahoney R, Pezzotti M and others. 2005. Molecular farming for new drugs and vaccines - Current perspectives on the production of pharmaceuticals in transgenic plants. EMBO Rep. 6(7):593-599.

Ma JKC, Drake PMW, Christou P. 2003. The production of recombinant pharmaceutical proteins in plants. Nat. Rev. Genet. 4(10):794-805.

Menkhaus TJ, Bai Y, Zhang CM, Nikolov ZL, Glatz CE. 2004a. Considerations for the recovery of recombinant proteins from plants. Biotechnol. Prog. 20(4):1001-1014.

Menkhaus TJ, Eriksson SU, Whitson PB, Glatz CE. 2002. Host selection as a downstream strategy: Polyelectrolyte precipitation of beta-glucuronidase from plant extracts. Biotechnol. Bioeng. 77(2):148-154.

Menkhaus TJ, Pate C, Krech A, Glatz CE. 2004b. Recombinant protein purification from pea. Biotechnol. Bioeng. 86(1):108-114.

Merle C, Perret S, Lacour T, Jonval V, Hudaverdian S, Garrone R, Ruggiero F, Theisen M. 2002. Hydroxylated human homotrimeric collagen I in Agrobacterium tumefaciens-mediated transient expression and in transgenic tobacco plant. FEBS Lett. 515(1-3):114-118.

Myllyharju J. 2005. Intracellular post-translational modifications of collagens. Top. Curr.Chem. 247:115-147.

Myllyharju J, Kivirikko KI. 2004. Collagens, modifying enzymes and their mutations in humans, flies and worms. Trends Genet. 20(1):33-43.

Myllyharju J, Lamberg A, Notbohm H, Fietzek PP, Pihlajaniemi T, Kivirikko KI. 1997. Expression of wild-type and modified pro alpha chains of human type I procollagen in insect cells leads to the formation of stable [alpha 1(I)](2)alpha 2(I) collagen

heterotrimers and [alpha 1(I)](3) homotrimers but not [alpha 2(I)](3) homotrimers. J. Biol. Chem. 272(35):21824-21830.

Myllyharju J, Nokelainen M, Vuorela A, Kivirikko KI. 2000. Expression of recombinant human type I-III collagens in the yeast *Pichia pastoris*. Biochem. Soc. Trans. 28:353-357.

Nandi S, Yalda D, Lu S, Nikolov Z, Misaki R, Fujiyama K, Huang N. 2005. Process development and economic evaluation of recombinant human lactoferrin expressed in rice grain. Transgenic Res. 14(3):237-249.

Nokelainen M, Lamberg A, Helaakoski T, Myllyharju J, Pihlajaniemi T, Kivirikko KI. 1996. Characterization of human type I, II and III collagens expressed in insect cells. Matrix Biol. 15(3):194-195.

Nokelainen M, Tu HM, Vuorela A, Notbohm H, Kivirikko KI, Myllyharju J. 2001. High-level production of human type I collagen in the yeast *Pichia pastoris*. Yeast 18(9):797-806.

Olsen D, Jiang J, Chang R, Duffy R, Sakaguchi M, Leigh S, Lundgard R, Ju J, Buschman F, Truong-Le V and others. 2005. Expression and characterization of a low molecular weight recombinant human gelatin: development of a substitute for animal-derived gelatin with superior features. Protein Expr. Purif. 40(2):346-357.

Olsen D, Yang CL, Bodo M, Chang R, Leigh S, Baez J, Carmichael D, Perala M, Hamalainen ER, Jarvinen M and others. 2003. Recombinant collagen and gelatin for drug delivery. Adv. Drug Delivery. Rev. 55(12):1547-1567.

Olsen DR, Leigh SD, Chang R, McMullin H, Ong W, Tai E, Chisholm G, Birk DE, Berg RA, Hitzeman RA and others. 2001. Production of human type I collagen in yeast reveals unexpected new insights into the molecular assembly of collagen trimers. J. Biol. Chem. 276(26):24038-24043.

Pakkanen O, Hamalainen ER, Kivirikko KI, Myllyharju J. 2003. Assembly of stable human type I and III collagen molecules from hydroxylated recombinant chains in the yeast *Pichia pastoris* - Effect of an engineered C-terminal oligomerization domain foldon. J. Biol. Chem. 278(34):32478-32483.

Pakkanen O, Pirskanen A, Myllyharju J. 2006. Selective expression of nonsecreted triple-helical and secreted single-chain recombinant collagen fragments in the yeast *Pichia pastoris*. J. Biotechnol. 123(2):248-256.

Parmenter DL, Boothe JG, vanRooijen GJH, Yeung EC, Moloney MM. 1995. Production of biologically active hirudin in plant seeds using oleosin partitioning. Plant Mol. Biol. 29(6):1167-1180.

Perret S, Merle C, Bernocco S, Berland P, Garrone R, Hulmes DJS, Theisen M, Ruggiero F. 2001. Unhydroxylated triple helical collagen I produced in transgenic plants provides new clues on the role of hydroxyproline in collagen folding and fibril formation. J. Biol. Chem. 276(47):43693-43698.

Przybycien TM, Pujar NS, Steele LM. 2004. Alternative bioseparation operations: life beyond packed-bed chromatography. Curr. Opin. Biotechnol. 15(5):469-478.

Ramessar K, Sabalza M, Capell T, Christou P. 2008. Maize plants: An ideal production platform for effective and safe molecular pharming. Plant Sci. 174(4):409-419.

Ritala A, Eskelin K, Holkeri H, Wahlström E, Baez J, Mäkinen K, Nuutila A. 2007. Development of a model system for the production of a recombinant industrial protein in transgenic barley grain. Annual Phytochemical Society of Europe.

Ruggiero F, Exposito JY, Bournat P, Gruber V, Perret S, Comte J, Olagnier B, Garrone R, Theisen M. 2000. Triple helix assembly and processing of human collagen produced in transgenic tobacco plants. FEBS Lett. 469(1):132-136.

Sakaguchi M, Hori H, Hattori S, Irie S, Imai A, Yanagida M, Miyazawa H, Toda M, Inouye S. 1999. IgE reactivity to alpha 1 and alpha 2 chains of bovine type I collagen in children with bovine gelatin allergy. J. Allergy Clin. Immunol. 104(3):695-699.

Samyn-Petit B, Gruber V, Flahaut C, Wajda-Dubos JP, Farrer S, Pons A, Desmaizieres G, Slomianny MC, Theisen M, Delannoy P. 2001. N-glycosylation potential of maize: The human lactoferrin used as a model. Glycoconj. J. 18(7):519-527.

Schillberg S, Fischer R, Emans N. 2003. Molecular farming of recombinant antibodies in plants. Cell. Mol. Life Sci. 60(3):433-445.

Schillberg S, Zimmermann S, Findlay K, Fischer R. 2000. Plasma membrane display of anti-viral single chain Fv fragments confers resistance to tobacco mosaic virus. Mol. Breed. 6(3):317-326.

Scopes RK. 1994. Protein purification: Principles and practice. Cantor CR, editor Springer Science+Business Media, Inc.

Sparrow PAC, Irwin JA, Dale PJ, Twyman RM, Ma JKC. 2007. Pharma-Planta: Road testing the developing regulatory guidelines for plant-made pharmaceuticals. Transgenic Res. 16(2):147-161.

Stoger E, Ma JKC, Fischer R, Christou P. 2005. Sowing the seeds of success: pharmaceutical proteins from plants. Curr. Opin. Biotechnol. 16(2):167-173.

Stoger E, Vaquero C, Torres E, Sack M, Nicholson L, Drossard J, Williams S, Keen D, Perrin Y, Christou P and others. 2000. Cereal crops as viable production and storage systems for pharmaceutical scFv antibodies. Plant Mol. Biol. 42(4):583-590.

Streatfield SJ. 2007. Approaches to achieve high-level heterologous protein production in plants. Plant Biotechnol. J. 5(1):2-15.

Streatfield SJ, Lane JR, Brooks CA, Barker DK, Poage ML, Mayor JM, Lamphear BJ, Drees CF, Jilka JM, Hood EE and others. 2003. Corn as a production system for human and animal vaccines. Vaccine 21(7-8):812-815.

Thommes J, Etzel M. 2007. Alternatives to chromatographic separations. Biotechnol. Prog. 23(1):42-45.

Toman PD, Pieper F, Sakai N, Karatzas C, Platenburg E, de Wit I, Samuel C, Dekker A, Daniels GA, Berg RA and others. 1999. Production of recombinant human type I procollagen homotrimer in the mammary gland of transgenic mice. Transgenic Res. 8(6):415-427.

Tomita M, Munetsuna H, Sato T, Adachi T, Hino R, Hayashi M, Shimizu K, Nakamura N, Tamura T, Yoshizato K. 2003. Transgenic silkworms produce recombinant human type III procollagen in cocoons. Nat. Biotechnol. 21(1):52-56.

Twyman RM, Stoger E, Schillberg S, Christou P, Fischer R. 2003. Molecular farming in plants: host systems and expression technology. Trends Biotechnol. 21(12):570-578.

Vaughan PR, Galanis M, Richards KM, Tebb TA, Ramshaw JAM, Werkmeister JA. 1998. Production of recombinant hydroxylated human type III collagen fragment in *Saccharomyces cerevisiae*. DNA Cell Biol. 17(6):511-518.

Vuorela A, Myllyharju J, Nissi R, Pihlajaniemi T, Kivirikko KI. 1997. Assembly of human prolyl 4-hydroxylase and type III collagen in the yeast *Pichia pastoris*: formation of a stable enzyme tetramer requires coexpression with collagen and assembly of a stable collagen requires coexpression with prolyl 4-hydroxylase. Embo J. 16(22):6702-6712.

Werten MWT, Van den Bosch TJ, Wind RD, Mooibroek H, De Wolf FA. 1999. High-yield secretion of recombinant gelatins by *Pichia pastoris*. Yeast 15(11):1087-1096.

Werten MWT, Wisselink WH, van den Bosch TJJ, de Bruin EC, de Wolf FA. 2001. Secreted production of a custom-designed, highly hydrophilic gelatin in *Pichia pastoris*. Protein Eng. 14(6):447-454.

Witcher DR, Hood EE, Peterson D, Bailey M, Bond D, Kusnadi A, Evangelista R, Nikolov Z, Wooge C, Mehigh R and others. 1998. Commercial production of beta-glucuronidase (GUS): A model system for the production of proteins in plants. Mol. Breed. 4(4):301-312.

Woodard SL, Mayor JM, Bailey MR, Barker DK, Love RT, Lane JR, Delaney DE, McComas-Wagner JM, Mallubhotla HD, Hood EE and others. 2003. Maize (Zea mays)-derived bovine trypsin: Characterization of the first large-scale, commercial protein product from transgenic plants. Biotechnol. Appl. Biochem. 38:123-130.

Xie RZ, Stretton S, Sweeney DF. 2001. Artificial cornea: Towards a synthetic onlay for correction of refractive error. Biosci. Rep. 21(4):513-536.

Yang C, Hillas P, Tang J, Balan J, Notbohm H, Polarek J. 2004a. Development of a recombinant human collagen-type III based hemostat. J. Biomed. Mater. Res. B 69B(1):18-24.

Yang CL, Hillas PJ, Baez JA, Nokelainen M, Balan J, Tang J, Spiro R, Polarek JW. 2004b. The application of recombinant human collagen in tissue engineering. BioDrugs 18(2):103-119.

Yildirim S, Fuentes RG, Evangelista R, Nikolov ZL. 2002. Fractionation of transgenic corn for recovery of recombinant enzymes. J. Am. Oil Chem. Soc. 79(8):809-814.

Zaman F, Kusnadi AR, Glatz CE. 1999. Strategies for recombinant protein recovery from canolaby precipitation. Biotechnol. Prog. 15(3):488-492.

Zhong GY, Peterson D, Delaney DE, Bailey M, Witcher DR, Register III JC, Bond D, Li CP, Marshall L, Kulisek E and others. 1999. Commercial production of aprotinin in transgenic maize seeds. Mol. Breed. 5(4):345-356.

Zhong QX, Xu L, Zhang C, Glatz CE. 2007. Purification of recombinant aprotinin from transgenic corn germ fraction using ion exchange and hydrophobic interaction chromatography. Appl. Microbiol. Biotechnol. 76(3):607-613.

CHAPTER 2. PURIFICATION AND CHARACTERIZATION OF A 44 kDa RECOMBINANT COLLAGEN I alpha 1 FRAGMENT FROM CORN SEED

A paper to be submitted to Journal of Agricultural and Food Chemistry

Cheng Zhang[1], Julio Baez[2], and Charles E. Glatz[3]

Abstract

Corn grain is currently used as feedstock for the production of many industrial products. If recombinant proteins with equivalency to proteins currently obtained from other sources could be produced in corn hosts as a co-product, there is great potential for lowering the cost of production of the protein. While several recombinant proteins have been produced in corn, the extent of their characterization has been limited and their compatibility with biorefinery production is poor. We have purified and characterized a potentially suitable co-product, a 44 kDa fragment of human collagen I alpha 1 (CIα1). The 44 kDa CIα1 fragment was selectively extracted from ground grain at low pH and purified with the additional steps of ion-exchange and gel filtration chromatography, resulting in 44 kDa CIα1 with >70% purity and 60% yield. The N-terminal sequence, amino acid composition and immunoreactivity of the purified corn grain-derived 44 kDa CIα1 closely matched that of an analogous 44 kDa CIα1 fragment produced by the yeast *Pichia*. The corn-derived 44 kDa CIα1fragment had an intact protein mass of 44,088 Da that is within 0.2% of the mass

[1] Primary researcher and author, responsible for purification process development and protein characterization
[2] Co-author, supplying information and materials on collagen/gelatin assays and chemistry
[3] Author for correspondence

calculated from the expected sequence. Tandem mass spectrometry confirmed the primary sequence of the corn grain-derived 44 kDa CIα1 with 78% coverage. The amino acid composition analysis indicated a low level of prolyl hydroxylation, as has been seen for other corn grain-derived recombinant proteins. Glycoprotein staining revealed no glycosylation. These results provided strong evidence that the corn grain-derived 44 kDa CIα1 fragment accumulated in corn grain with the intended composition and post-translational modifications. This is the first expression of a collagenous protein in a cereal seed.

Key words: recombinant collagen fragment, gelatin, protein extraction, purification, characterization, chromatography, transgenic corn, maize.

Introduction

Though several safety and social concerns remain to be addressed, use of corn as an expression host has attractive potential for intermediate-value, recombinant protein-based products, particularly if the expression and post-translational processing can give an equivalent or even functionally superior product. Intermediate-valued products can be economically viable because corn grain biorefining allows the generation of several products in addition to the protein stream, such as starch and oil that are converted into diverse industrial products. In 2007 a record of 92 million acres of corn were planted in the United Sates with an expected yield of over 350 million tons of grain (*1*). About 75% of the crop was available for domestic animal feed and industrial use in 2007 (*2*). There is extensive experience with the cultivation and handling of genetically modified corn in the United

States. as in 2007 over 73% of the U.S. corn crop was planted with transgenic plants designed for herbicide and insect tolerance (*3*). Hence, there is no shortage of potential cost-effective co-production of so-called "Plant Made Industrial Proteins" (PMIP) required at large volumes (thousands of metric tons per year) and low cost (less than US$100/kg). Examples of potential PMIP applications are drug delivery (e.g. gelatin for capsules), bioprocessing (e.g. cellulases, lignases, xylanases, proteases, amylases, lipases, isomerases for feedstock conversion to biofuels), consumer products (e.g. detergent enzymes, fabric-treatment enzymes, bleaching enzymes, antifreeze peptides), antibiotic peptides, hazardous biological agent deactivating proteins (e.g. butyrylcholinesterase as a protective agent against organophosphate poisoning), and advanced performance protein-based polymers (e.g. silk, elastin).

We selected one such protein, a recombinant collagen fragment, potentially compatible with high-volume, economical co-production in a corn-based biorefinery. Recombinant collagen fragments generated from yeast fermentation have been used as substitutes for animal-derived gelatin, a denatured collagen-derived industrial protein used extensively in food, photography, and other pharmaceutical applications (*4, 5*). Over 50,000 metric tons of gelatins are consumed every year just for pharmaceutical use, with over 80% of this material used to manufacture capsules for the oral delivery of drugs and nutraceuticals. Use of plant-derived products, especially in pharmaceutical applications, could avoid the allergenicity and safety concerns related to animal-derived contaminants, lack of traceability, and inconsistent product quality due to the variability of tissue origin and animal age.

The currently preferred recombinant technology for gelatin production is based on the yeast *Pichia* fermentation (*5*). The recombinant *Pichia* production system has been

developed to produce many recombinant collagen fragments (from 8 to 100 kDa), as well as fully prolyl hydroxylated triple-helical recombinant collagen when paired with the co-expression of the human enzyme prolyl 4-hydroxylase (P4H) (*4-6*). Studies using a series of the *Pichia*-derived recombinant collagen I α1 (rCIα1) fragments ranging in size from 56 to 1014 amino acids indicated that these recombinant collagen fragments can be used as replacement for animal-derived gelatin for many of the current medical applications (*5*). CIα1 is the most common type of collagen found in animal-derived gelatin. The rCIα1 fragments are secreted in *Pichia* fermentation as fully intact species as determined by electrophoresis gels, N-terminal sequencing, gel filtration chromatography and mass spectrometry. The accumulation of recombinant collagen and of collagen fragments has also been demonstrated in other recombinant expression systems such as mammalian cells (*7-9*), insect cell culture(*10, 11*), *Escherichia coli* (*12, 13*), *Bacillus* (*14*), silkworms (*15*), milk of transgenic animals (*16, 17*) and other yeasts (*Sacchromyces cerevisia*e (*18, 19*), *Hansenula polymorpha* (*20, 21*)).

The degree of hydroxylation provided during expression is important to the functional properties of collagen because it modulates gelation and triple-helix formation (*5, 22*). Partially prolyl hydroxylated recombinant collagen was produced in tobacco cell cultures co-expressing a cloned animal P4H enzyme (*5*) and tobacco plants (*23, 24*). Transgenic tobacco cell cultures accumulated recombinant type-III collagen co-expressed with human P4H in a triple-helical conformation with 75% of the prolyl hydroxylation level found in collagen obtained from human tissue (*5*). Lack of prolyl hydroxylation when CIα1 accumulated in tobacco without cloned animal P4H suggested that any plant endogenous P4H is not able to hydroxylate proline at a detectable level (*24*). Recent studies indicated that transient and

stable expression of CIα1 can be achieved in barley and rice suspension cell culture (*25, 26*) and in barley and corn seeds (*27, 28*).

Consistently high accumulation levels (10-40 kg recombinant protein/acre, 2.5-10 g /kg grain) using tissue-specific promoters in high yielding commercial lines producing up to 4000 kg grain/acre will be required for the commercialization of recombinant PMIP. This high level of recombinant PMIP accumulation has been achieved after extensive selection and crossing. For example, the accumulation level for a recombinant avidin was increased 150-fold to 10 g avidin/kg grain in eight generations compared with the accumulation seen in the first transgenic grain (*29-36*). The early generation transgenic plants usually have very low levels of recombinant protein accumulation for characterizing the expressed protein. Thus, this report describes the development of a recovery and purification process for a recombinant collagen fragment derived from corn grain with low accumulation level (20 mg/kg) typical from the initial field trials conducted in the early stages of the development program. It is at this stage that characterization will be important because it could reveal molecular changes as produced in the seed or as may occur during grain handling and processing.

The collagen extraction was patterned after studies showing that pH and ionic strength have the greatest effects on protein extraction from grain. The total extractable corn endogenous protein concentration increased 10-fold when increasing the pH from 3 to 10 (*29, 37*). The total extractable corn endogenous protein concentration also increased with addition of NaCl up to 300 mM (*29*). Collagen fragments are acid soluble (*24*) and *Pichia*-derived CIα1 showed maximum solubility at pH 1-2, offering the possibility for selective extraction of CIα1 at low pH. *Pichia*-derived gelatin (*6*) and tobacco-derived collagen (*38*) were

purified using ion-exchange chromatography (IEC) eluted with acetate buffer. Gel filtration chromatography should be able to remove the remaining corn grain-related impurities as most of the extractable corn endogenous proteins at acidic pH are expected to be smaller than 30 kDa (*39*).

The characterization of the corn–derived 44 kDa CIα1 was based on the analysis of its immunoreactivity to CIα1-specific antibodies, amino acid composition determination, glycosylation staining, and mass spectrometry and peptide mapping to assess the fidelity of translation and post-translational modifications. Analytical data of the corn grain-derived CIα1 were compared with data from the *Pichia*-derived analogous CIα1 fragment. Previous publications of corn grain-derived recombinant proteins have focused on the production and downstream processing, providing only limited protein characterization data for non-glycosylated 6.5kDa aprotinin (*29*), β-glucuronidase (GUS), and avidin (*34*). The most extensive characterization was for corn-derived trypsin, for which Western blotting, MS, and glycosylation staining were used (*35*). All analytical results of the corn seed-derived trypsin were equivalent to the native protein with the exception of the molecular weight of the extracted protein. The differences in molecular weight were attributed to O-linked glycosylation of the corn-derived protein, demonstrated by the molecular weight after chemical deglycosylation. Karnoup et al. (*40*) showed that O-linked glycosylation found in a corn grain-derived human IgA was through hydroxylated proline residues indicating that a maize prolyl hydroxylase was able to convert proline to hydroxyproline and subsequently glycosylate IgG1. The cloned CIα1 fragment does not contain the consensus sequences for N-linked glycosylation. The only possibility for glycosylation is O-linked glycosylation, which could occur through serine, threonine or any hydroxylated proline or lysine residues.

Materials and Methods

Materials

The transgenic corn grain expressing the 44 kDa CIα1 was provided by the former ProdiGene Inc. (College Station, TX). This grain was harvested in the fall of 2004 from a field trial in Nebraska growing second-generation transgenic plants containing the gene for the expression of a 44 kDa collagen fragment of the triple-helical region of human CIα1. The CIα1 fragment has 498 amino acids resulting in a calculated MW of 44,002.8 Da, assuming proper cleavage of the signal sequence; its estimated pI is 8.1. Upon receipt at Iowa State University (ISU), the seeds were stored at 4^{o}C. A lyophilized 44 kDa CIα1 of the same sequence produced in *Pichia pastoris* (*Pichia*-derived 44 kDa CIα1) was provided by FibroGen Inc. (South San Francisco, CA) to be used as reference material.

Vector Construction and Transformation of Corn-derived 44 kDa CIα1

ProdiGene used the corn expression vector PGN9004 to clone the gene for the 44 kDa CIα1 fragment between the Hind III and Not I sites of the vector. The expression of the 44 kDa CIα1 gene was regulated by an embryo-specific maize globulin-1 promoter (PGNpr2) (*41*). The pin-II transcription terminator from the protease inhibitor II gene from potato (*42*) was used in this construct. For transgenic calli selection the phosphinothricin acetyl transferase gene was included in the vector under the control of the CaMV 35S promoter. This gene confers resistance to the herbicide bialaphos (*43-45*). The expression vector provided a signal sequence - barley alpha amylase signal sequence - to direct 44 kDa CIα1 to the cell wall.

The resultant vector was introduced into corn immature embryos using *Agrobacterium* transformation (*46*). The Hi-II maize line was used as host which initiates Type II embryogenic callus in culture selected using bialaphos (*47*). The use of the Hi-II calli growth medium (*47*) results in redifferentiation of the plant cells and regeneration into a plant allowing the scutellum cells to become embryogenic callus.

Embryos were cultured to recover 20 independent transgenic events and regenerated to recover transgenic first-generation (T0) plants. The first generation seeds (T1) produced from the T0 plants were transferred to a greenhouse to produce T1 plants that were pollinated using pollen donors from elite high yielding inbreds lines. The Hi-II maize line that is used in tissue culture for plant transformation has poor agronomic characteristics and is not high-yielding in the field. The T2 seeds were collected and planted in the field trial generating the grain (T3) used for this study.

Corn Grain Fractionation and Defatting

The 44 kDa CIα1-containing corn grain was separated into germ-rich and endosperm-rich fractions by using a dry milling procedure established at the Center for Crops Utilization Research (CCUR) at ISU (*48*). The germ-rich fraction was milled into flour using a household coffee mill (GE, model 169028) and then defatted. Defatting was carried out by mixing the ground germ-rich corn flour with hexane (1:5 w/v) using a magnetic stir bar for 60 min at 0^{o}C. The supernatant was decanted after centrifugation (15 min, 2000*g*, 23^{o}C) and the flour then extracted a second time as above. The residual hexane in the flour was removed by air drying at room temperature.

Recombinant 44 kDa CIα1 Recovery

The extraction was designed to recover the corn grain-derived 44 kDa CIα1 with minimal co-extraction of the endogenous corn grain components thus reducing the complexity of the extract for the subsequent chromatographic steps. The defatted germ-rich corn flour was extracted twice by mixing with extraction buffer (0.1M phosphoric acid, 0.15 M sodium chloride, pH 1.8) in a ratio of 1:5 w/v for 3 h at room temperature. The extracts were clarified by centrifugation (~3000*g*, 10 min at 23^{o}C), then pooled and dialyzed (Pierce SnakeSkin T Dialysis Tubing, 3.5K MWCO) overnight at 4^{o}C against sodium acetate buffer (50 mM, pH 4.6). The dialyzed extract was filtered through a 0.45 μm membrane (GV Durapore PVDF, Millipore) before chromatography.

Recombinant 44 kDa CIα1 Purification

The dialyzed extract was fractionated by cation-exchange chromatography (CEC) using a XK16/20 column (GE Healthcare) packed with 12 ml of SP Sepharose^{TM} Fast Flow resin (Sigma) equilibrated with 50 mM sodium acetate (pH 4.6). The elution was carried out over 15 column volumes using a salt concentration gradient from 0 to 0.5 M sodium chloride in 50 mM sodium acetate (pH 4.6) at 3 ml/min flow rate. The collected fraction with highest 44 kDa CIα1 content as determined by ELISA was concentrated 5X using a Centriplus YM3 (Amicon, 3K MWCO) filter before proceeding to the next chromatographic step. The concentrated CEC pool was further purified by gel filtration chromatography (GFC) using a XK 16/40 column (GE Healthcare) packed with 70 ml of Sephacryl S-200HR GFC medium (GE Healthcare). Protein elution was carried out with a 50 mM sodium acetate, 0.15 M

sodium chloride (pH 4.6) buffer at a flow rate of 0.5 ml/min. The two fractions with highest 44 kDa CIα1 content as determined by ELISA were used for the characterization studies.

Total Soluble Protein Assay

Total soluble protein concentration was determined using the Coomassie Plus - The Better Bradford™ Assay Kit (Pierce). Bovine serum albumin (BSA) was used as standard. Protein concentration was the average of three replicate assays.

Enzyme-Linked ImmunoSorbent Assay (ELISA)

A competitive ELISA was used to determine the 44 kDa CIα1 concentration. A 96-well plate (Corning Costar High Binding Cat# 3590) was coated with 2 μg/ml of streptavidin (Zymed #43-4301) in PBS (Mediatech #21-040-CV) and incubated at 4^{o}C overnight. Biotinylated *Pichia*-derived 44 kDa CIα1 (125 ng/ml in assay buffer (100 mM PBS, 0.05% Tween 20, 0.1% BSA, 0.1% Kathon CG/ICP, pH 7.0)) was then immobilized to the plate by incubation with the streptavidin at 4^{o}C overnight and worked as a competitor against corn grain-derived 44 kDa CIα1 for the primary antibody. The primary antibody was a rabbit polyclonal antibody (FibroGen), raised against a *Pichia*-derived 25kDa CIα1, and used at 1:4000 dilution. The 25kDa CIα1 was part of the 44 kDa CIα1, and this antibody was reactive with both. Horseradish peroxidase (HRP) –conjugated goat anti-rabbit IgG (H+L) (Invitrogen) was used as the secondary antibody at a 1:5000 dilution. Both primary and secondary antibodies were diluted in the assay buffer. A SureBlue^{TM} TMB substrate solution (Kirkegarrd & Perry Laboratories) was used as the substrate for the HRP conjugate and incubated with it at ambient temperature (23^{o}C) for 30 min. The plate was then read at 450

nm on a microplate reader (EL340, BioTek Instruments). *Pichia*-derived 44 kDa CIα1 of known concentration was the standard for the quantification of the corn grain-derived 44 kDa CIα1.

SDS-PAGE, Western Analysis, and Glycosylation Detection

SDS-PAGE was carried out by following the manual of the Ready Gel® Precast Gels Application Guide (BioRad) using a Mini-PROTEAN® II Electrophoresis Cell (BioRad). Protein samples were diluted in the sample buffer described in the Application Guide at a 1:1 ratio and heated at 100°C for 5 min before being loaded to a 4-15% Tris-HCl precast gel (BioRad). The *Pichia* 44 kDa CIα1 was used as reference material.

For Western analysis, the protein bands from the gel were transferred to a polyvinylidene difluoride (PVDF) membrane (Invitrogen) after SDS-PAGE. The membrane was incubated with a rabbit polyclonal antibody (primary antibody with 1:4000 dilution) overnight at 4°C, followed by incubation with horseradish peroxidase (HRP) –conjugated goat anti-rabbit IgG (H+L) (Invitrogen) (secondary antibody with 1:5000 dilution) for 1 h at 23°C. An ECL plus Western blotting detection system (Amersham Biosciences, RPN 2132) was used to visualize the immunoreactive protein bands. The membrane was exposed to a Fuji Super Rx film (Fisher Scientific, 0444199) for about 30 sec and the film was subsequently developed in a Futura™ 2000K automatic X-ray film processor (Fischer Industries, Inc.).

A Gelcode Glycoprotein Staining Kit (Pierce) was used to detect the presence of glycosylation in the purified corn 44 kDa CIα1. After SDS-PAGE, the gel is stained with the kit reagents resulting in only the protein bands containing glycoproteins developing a

magenta color. A replicate gel was stained with Coomassie Brilliant Blue to visualize all the proteins.

Intact Protein Mass Determination by Mass Spectrometry (MS)

MALDI-TOF (Matrix-Assisted Laser Desorption Ionization-Time of Flight) MS was used to measure the mass of the intact protein. The GFC pooled fractions with the highest 44 kDa CIα1 content were further purified to remove impurities by reverse phase high-pressure liquid chromatography (RP-HPLC) (Phenomenex, Jupiter C4, 5µm, 300Å, 250 x 4.6 mm) before the MS analysis. The elution was carried out using a linear gradient of 30-60 % acetonitrile in 0.1% trifluoroacetic acid over 30 min at 1 ml/min flow rate. The isolated 44 kDa CIα1 was freeze-dried, reconstituted with DI water, and then mixed with 3, 5-dimethoxy-4-hydroxy-cinnamic acid (sinapinic acid) used as the matrix. The mixture was spotted on a MALDI sample plate and air-dried. The external calibration was performed using bovine serum albumin. The analysis was done using a Voyager-DE STR MALDI-TOF MS (PerSeptive Biosystems) in the Protein Facility at ISU.

Internal Peptide Sequencing by Tandem MS

Two types of proteolytic digestions were performed for individual analysis by MS. Sequencing-grade trypsin (Promega) was added to the corn 44 kDa CIα1 at an enzyme to substrate ratio of 1:20 (w/w) in 50 mM ammonium bicarbonate buffer, pH 7.8 with the digestion carried out for 3 h at 37^{o}C. Sequencing grade endoproteinase Glu-C (New England Biolabs) was added to the 44 kDa CIα1 at an enzyme to substrate ratio of 1:20 (w/w) in 50 mM Tris-HCl, 0.5mM Glu-Glu, pH 8.0 (New England Biolabs) with the digestion carried out

for 18 h at 23^{o}C. Both digests were diluted 2X with 50% methanol and 0.5% formic acid, and the resultant solutions were further purified and concentrated using a C18 ZipTip pipette tips (Millipore). About 20 µl of each solution were loaded separately to the ESI-quadrupole-TOF tandem mass spectrometry (Qstar XL, Applied Biosystems). Analysis of the MS and tandem MS (MS/MS) spectra data was done using the Mascot computer algorithm which allows for peptide amino acid sequence interpretation. A self-defined database with the human CIα1 amino acid sequence containing the expected sequence for the 44 kDa CIα1 fragment was selected and used for the analysis. Both MS spectra collection and amino acid sequence interpretation were carried out in Proteomics Facility at ISU.

N-terminal Sequencing

The purified corn grain-derived 44 kDa CIα1 was further separated from minor impurities by SDS-PAGE, and transferred to a PVDF membrane. The 44 kDa region band was cut out and loaded onto the Procise protein sequencer (Applied Biosystems). Cycles of Edman degradation were conducted followed by injection onto RP-HPLC. The first 20 amino acids on the N-terminus were sequenced. The sequencing was performed at the ISU Protein Facility with confirming results provided by FibroGen Inc.

Amino Acid Composition Analysis

For the amino acid composition analysis, the PVDF membrane excised 44 kDa CIα1 band from SDS-PAGE was subjected to hydrolysis under vacuum (6 N HCl, 150^{o}C, 65 min). The resultant amino acids were then derivatized with phenylisothiocyanate (PTC) by using a PTC derivatizer (Perkin Elmer Applied Biosystems Model 420A), and analyzed by a PTC

Amino Acid Analyzer (Perkin Elmer Applied Biosystems Model 130A). A *Pichia*-derived 44 kDa CIα1 (in liquid solution instead of a PVDF extracted band) was also submitted and analyzed for comparison. Trans-4-hydroxy-L-proline (Sigma H5534) was used as standard for quantifying the hydroxyproline content. The amino acid composition of the 44 kDa CIα1 samples was performed at the ISU Protein Facility.

Results and Discussion

A summary of the results of the recovery and purification process is provided in Table 1.

Protein Extraction

The total soluble protein (TSP) extracted (Table 1) from the 5 g of corn germ-rich flour represents less than 10% of the maximum TSP in corn germ due to the low pH of the extraction buffer. This value is similar to those obtained by other researchers using acidic pH for extraction (*37*). Recovery of 44 kDa CIα1 from the enriched germ fraction was calculated at 120 mg 44 kDa CIα1/kg corn germ fraction at a 2% purity based on TSP in this acidic extract. Analysis of the grain before fractionation to obtain the enriched germ used for the extraction indicated a concentration of 20 mg 44 kDa CIα1/kg corn grain consistent with values seen in the seed used for planting the field trial. To achieve commercially acceptable levels it will be required to increase the CIα1 accumulation level in corn grain at least 50-to 250-fold to achieve 1 to 5 g of 44 kDa CIα1/kg corn grain. With a similar system ProdiGene achieved levels for a recombinant avidin of 10 g/kg grain that was 150-fold higher than the accumulation seen in the first transgenic grain (T1) (*32*).

Table 1. Purification of the corn grain-derived 44 kDa CIα1

Step	Volume (ml)	Protein (µg)	44 kDaCIα1 (µg)	Purity by ELISA* (%)	Purity by gel** (%)	Step purification factor***	Yield (%)
Extract	67	20,971	433	2.06	1.2	1	100
IEC	8	1,370	316	23.10	10	11.2	73
GFC load	0.9	1,119	259	23.10	10	—	—
pooled	10	76.5	213	100	70	4.3	82

* Purity by ELISA was based on the amount of 44 kDa CIα1 as determined by ELISA and total protein as determined by Coomassie blue protein assay using BSA as a reference standard. The difference between the sample and reference proteins leads to overestimates of purity when the sample purity was high and concentration was low because collagen gives a low reading in this assay.

** Purity was estimated by the densitometric analysis of SDS polyacrylamide gel stained with Coomassie Brilliant Blue where dye-binding by collagen would result in low estimates of purity.

*** Step purification factor was determined by the purities calculated with ELISA, and overall purification factor was 48.5.

Corn-derived 44 kDa CIα1 Purification

Figure 1 shows the elution profile of the 44 kDa CIα1 from the first chromatographic step, the cation-exchange column. No significant amount of the 44 kDa CIα1 was detected during the sample loading and washing steps, indicating that the 44 kDa CIα1 was effectively captured by the column. The 44 kDa CIα1 was eluted within the first main peak, as shown by the black arrow in Figure 1, and collected mainly in three fractions. One of these fractions was an order of magnitude richer in 44 kDa CIα1 than the other two and it was carried forward for the second chromatographic step. This fraction was 11X-enriched, 23% pure, and contained 73% of the 44 kDa CIα1 loaded to the column (Table 1).

Figure 2 shows the elution profile of the 44 kDa CIα1 from the second chromatographic step (GFC). The 44 kDa CIα1 eluted mainly over two fractions before the elution of the majority of the native corn proteins as shown by the arrow in Figure 2. The yield of 44 kDa CIα1 for this purification step was 82% with a purity of 100% based on the ratio of 44 kDa CIα1 content to the total protein content in the pooled fractions. This purity is based on ELISA for the CIα1 fragment and the total protein assay by using BSA as a standard. An alternative estimate based on densitometry using the SDS-PAGE gels (Figure 3, lane 4) yielded a purity value of 70%. The discrepancy between the two purity results stems from the low binding of the dye by collagen in the two assays.

SDS-PAGE was used to monitor the purification process performance (Figure 3). The extract (Lane 2) mainly consists of proteins smaller than 25 kDa, assuming that there are not many multimeric proteins. The 44 kDa CIα1 was partially enriched after the cation-exchange step (Lane 3), and further purified after the GFC step (Lane 4). While some native

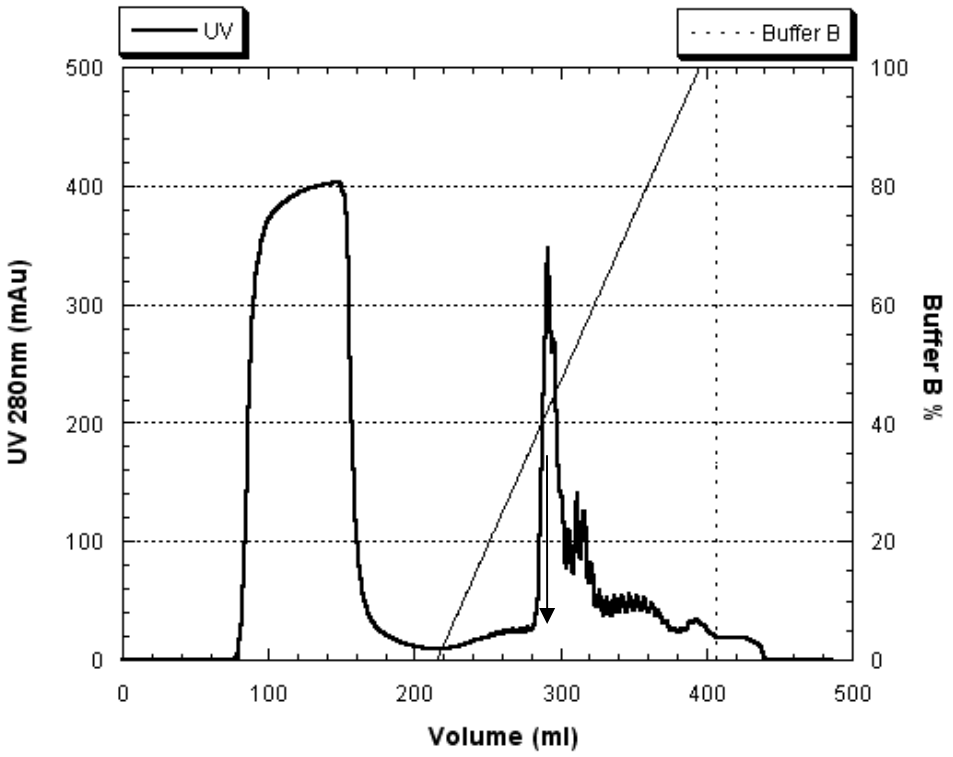

Figure 1. Elution profile of corn grain-derived 44 kDa CIα1 from SP-Sepharose® Fast Flow cation exchange chromatography. The elution was carried out over 15 column volumes of a 0 to 0.5 M sodium chloride gradient in 50 mM sodium acetate (pH 4.6) at 3 ml/min. The arrow shows the region where 44 kDa CIα1 was eluted. (Buffer B: 0.5 M sodium chloride + 50 mM sodium acetate).

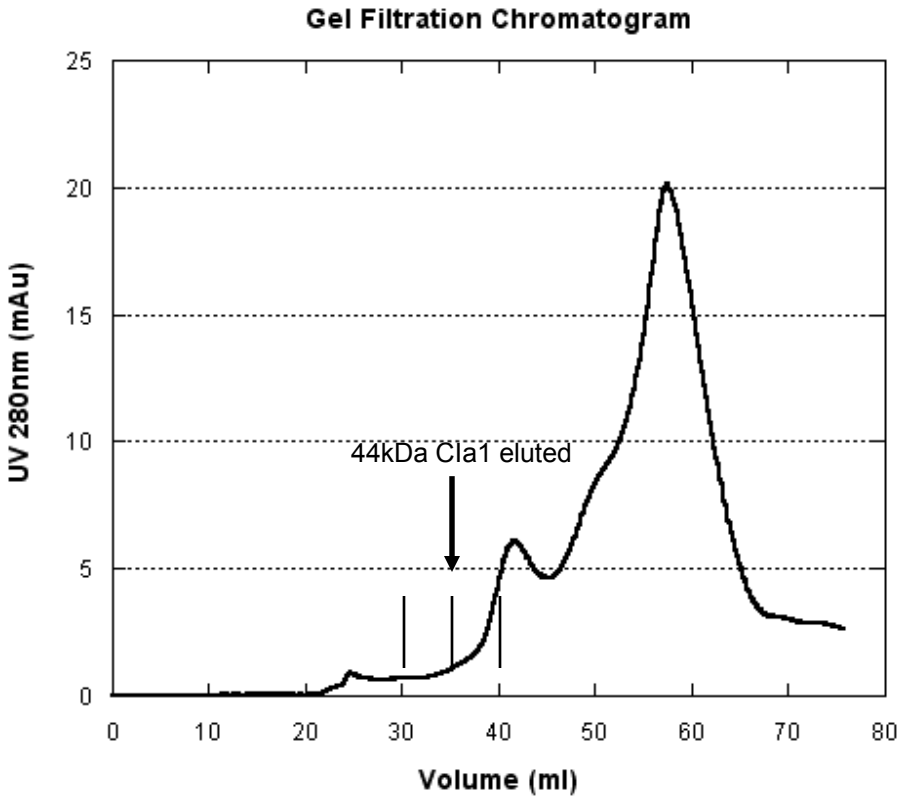

Figure 2. Elution profile of corn grain-derived 44 kDa CIα1 from Sephacryl S-200HR gel filtration chromatography. Elution was carried out with 50 mM sodium acetate plus 0.15 M sodium chloride (pH 4.6) at 0.5 ml/min. The arrow indicates the region where 44 kDa CIα1 was eluted. Two fractions, each with 5 ml/fraction, were collected.

corn proteins with MW similar to the 44 kDa CIα1 remained in the GFC product fraction, most small corn proteins were removed during this step. The *Pichia*-derived 44 kDa CIα1 (first band in Lane 5) is shown for comparison and shows that the corn-derived 44 kDa CIα1 had a very similar MW to its *Pichia* counterpart. The MWs of both corn and *Pichia*-derived 44 kDa CIα1 estimated from the gel were about 60 kDa, 36% higher than the expected MW of 44 kDa. A similar phenomenon was reported by Werten et al.(*49*) when analyzing a 36.8 kDa recombinant collagen fragment produced from *Pichia pastoris*. Their explanation was that the collagen fragment's high hydrophilicity resulted in low binding to the SDS, leading to slower migration.

Immunorecognition

Western analysis demonstrated the immunoreactivity of the corn grain-derived 44 kDa CIα1 (Figure 4). The *Pichia*-derived 44 kDa CIα1 was also included in this analysis for comparison. The results showed that the corn grain-derived CIα1 had a very similar molecular weight to its *Pichia* counterpart, and that both the corn and the *Pichia*- derived 44 kDa CIα1s were reactive to the same 25kDa CIα1 specific antibody confirming the reactivity seen in the ELISA. A faster migrating corn grain-derived 44 kDa CIα1 immunoreactive protein band was observed when the sample loading to the gel is higher (lane 4, Figure 4). Similar phenomena were reported by others producing CIα1 fragments. Werten et al (*50*) reported that CIα1 fragment produced by *Pichia pastoris* was susceptible to proteolytic degradation. Other heterologous proteins expressed in corn had also shown susceptibility to proteolytic degradation. A rootworm-control recombinant protein expressed in genetically modified corn plants exhibited a proteolytic fragment with MW slightly smaller than the

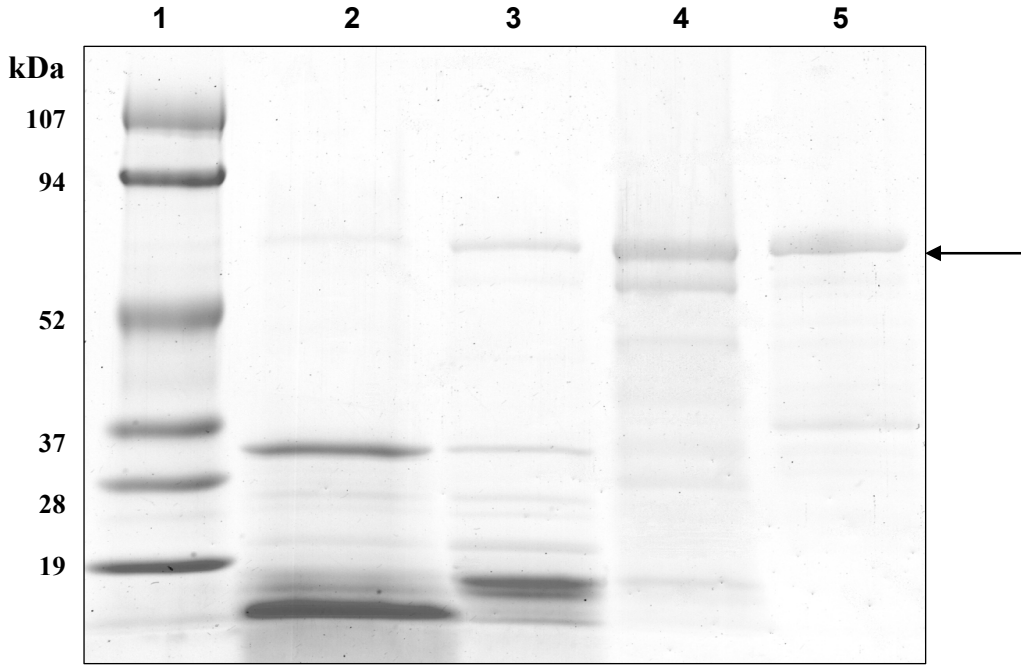

Figure 3. Coomassie Brilliant Blue -stained SDS-PAGE: Lane 1, protein marker; Lane 2, corn grain-derived 44 kDa CIα1 extract; Lane 3, IEC fraction; Lane 4, GFC pooled fractions; and Lane 5, *Pichia*-derived 44 kDa CIα1. The arrow indicates the position of the 44 kDa CIα1.

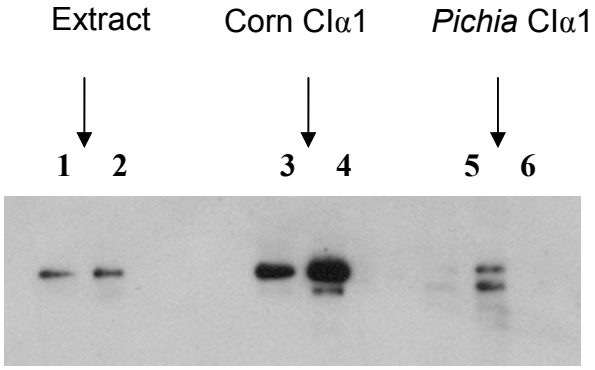

Figure 4. Western blotting: Lane 1and 2, corn grain-derived 44 kDa CIα1 extract; Lane 3 and 4, purified corn grain-derived 44 kDa CIα1; Lane 5 and 6, Pichia-derived 44 kDa CIα1 supplied by FibroGen. For all the lanes above, the two lanes next to each other are loaded with different amounts of the same materials.

target protein in Western blotting (*51*). Therefore it is reasonable to suspect that the observed faster migrating protein is a proteolytic fragment resulting from of the activity of corn grain-derived endogenous proteases during crop growth, grain storage or protein extraction. The proteolytic fragment was not observable until the protein was purified and enriched at the later stage of the purification process (lane 4, Figure 3).

Intact Protein Mass Determination by MS

The MW of the intact corn grain-derived 44 kDa CIα1 was measured by MS. The measured intact protein mass of the corn grain-derived 44 kDa CIα1 was estimated from MS to be 44,088 Da, which is within 0.2% of the mass (44002 Da) calculated from the amino acid sequence expected from the gene used for transformation. The accuracy for this MS analysis as provided by the manufacturer is ±0.05% for a 16 kDa protein. However, for proteins with larger size, accuracy would be worse. Adams et al., (*52*) reported an average of 0.43% difference between measured intact mass by MS and the expected mass when using the same type of MALDI-TOF to analyze zein proteins (ranging from 10-30 kDa) in corn kernel. Therefore the 0.2% mass difference in this work is estimated to be within the accuracy range of the equipment and should not be considered as evidence for protein modification. The intact protein mass of the 44 kDa CIα1 provided reliable evidence that the corn-derived 44 kDa CIα1 was correctly expressed in the corn without significant post-translational modification.

Peptide Sequencing by Tandem MS

Tandem MS can be used to verify that a particular protein is being produced with as little as two to three peptides corresponding to 10-15% sequence coverage (*53*) on the basis of the low probability of another protein containing the same combination of peptides. Complete sequence coverage is recommended to identify all possible post-translational modifications on amino acid residues. To yield high overlapping sequence coverage for 44 kDa CIα1, two proteolytic enzymes, trypsin and endoproteinase Glu-C, were used to perform digestions. Trypsin predominantly cleaves proteins at the carboxyl side of lysine and arginine, except when either is followed by proline, while Glu-C cleaves peptide bonds C-terminal to glutamic acid residues. Theoretical analysis of these digestions for the expected amino acid sequences with the two enzymes was performed by an online program (Protein Prospector http://pospector.ucsf.edu) predicting complementary coverage for the 44 kDa CIα1. Sequence coverage of the corn grain-derived 44 kDa CIα1 resolved by tandem MS with both trypsin and Glu-C digestion is shown in Figure 5. Less sequence coverage was obtained from the Glu-C digest (28%) than from the trypsin digest (64%). Glu-C digestion generated longer peptides since it had fewer cleavage sites on the 44 kDa CIα1 than trypsin. The longer peptides often contained more basic amino acid residues, which could lead to the formation of ions with higher charge state. The highly charged peptides are usually more difficult to fragment further in generating the MS/MS spectrum data (*54*), thus resulting in reduced sequence coverage. The large 54-amino acid tryptic fragment from 454G (glycine) to 507P (proline) in Figure 5, which was not resolved by tandem MS, contained only two potential cleavage sites for trypsin. The theoretical digestion of the fragment with Glu-C by the program Protein Prospector yielded ten short fragments, ranging from 11-23 amino acids. The range represents an ideal peptide length for MS fragmentation and analysis (*53*). The

actual digestion of the fragment with Glu-C yielded four peptides, from which a 54-amino acid fragment was totally resolved as shown in bold italics of Figure 5.

It should be noted that the MS-MS data could not rule out the possibility of errors in the unmatched sequences. Limited instrument sensitivity and incomplete MS spectra make it extremely difficult to obtain 100% coverage (*55*). However, the 79% sequence coverage confirmed by tandem MS and N-terminal sequencing and the agreement of intact MW obtained in this work provide confidence that the 44 kDa CIα1 produced from the corn was consistent with the gene used for the expression with minimal post-translation modification.

```
351      GLPGAK GLTGSPGSPG PDGKTGPPGP AGQDGRPGPP GPPGARGQAG
401 VMGFPGPKGA AGEPGKAGER GVPGPPGAVG PAGKDGEAGA QGPPGPAGPA
451 GERGEQGPAG SPGFQGLPGP AGPPGEAGKP GEQGVPGDLG APGPSGARGE
501 RGFPGERGVQ GPPGPAGPRG ANGAPGNDGA KGDAGAPGAP GSQGAPGLQG
551 MPGERGAAGL PGPKGDRGDA GPKGADGSPG KDGVRGLTGP IGPPGPAGAP
601 GDKGESGPSG PAGPTGARGA PGDRGEPGPP GPAGFAGPPG ADGQPGAKGE
651 PGDAGAKGDA GPPGPAGPAG PPGPIGNVGA PGAKGARGSA GPPGATGFPG
701 AAGRVGPPGP SGNAGPPGPP GPAGKEGGKG PRGETGPAGR PGEVGPPGPP
751 GPAGEKGSPG ADGPAGAPGT PGPQGIAGQR GVVGLPGQRG ERGFPGLPGP
801 SGEPGKQGPS GASGERGPPG PMGPPGLAGP PGESGREGAP GAEGSPGRDG
851 SP
```

Figure 5. Combined sequence coverage 79% resolved by tandem MS with both trypsin and Glu-C digestion, and by N-terminal sequencing. The 44 kDa CIα1 starts at 355G, and ends at 852P. Peptides in bold are derived from trypsin digestion (64% coverage), peptides in bold italic are from Glu-C digestion (28% coverage), and the peptides in bold and underline are the overlapped sequences.

N-terminal Sequencing

The first 20 amino acid residues at the N-terminus, "GLPGAKGLTGSPGSPGPDGK", all match that expected from the gene sequence, confirming the identity of the 44 kDa CIα1 and indicating that the signal sequence was properly cleaved.

Amino Acid Analysis

Amino acid analysis (Table 2) showed that the corn grain-derived 44 kDa CIα1 was similar to its *Pichia* counterpart for all the amino acids measured except in the proline and hydroxyproline content. While *Pichia* is not able to hydroxylate proline (*5*), corn seed seems to be capable of hydroxylating prolines though to a limited degree compared with the level obtained in human collagen . The 2% hydroxyproline content in the corn grain-derived 44 kDa CIα1 suggested the presence of an endogenous prolyl hydroxylase in corn seed and that mammalian prolyl 4-hydroxylase (P4H) co-expression will be necessary to obtain higher levels of hydroxylation in corn seed. The accumulation of hydroxylated CIα1 fragments has been reported earlier in both constitutive, glucose-fed fermentations and methanol-induced yeast *Hansenula polymorpha* fermentations, but only when a complex nutrient source (peptone) was added to the media (*21*). No hydroxylation was observed when the fermentation was conducted with a medium containing only salts and methanol (*56*). In contrast, when the same fragment was expressed in *Pichia* in the presence of peptone, no hydroxylation occurred. Hydroxyproline was also absent in all other CIα1 fragments expressed in *Pichia*, indicating that the presence of an endogenous prolyl hydroxylase activity was specific to *Hansenula* (*21*).

Table 2. Amino acid composition of the corn grain-derived 44 kDa CIα1

Amino acid	*Pichia* 44 kDa CIα1 (%)	Corn 44 kDa CIα1 (%)	Human type I homotrimer * (%)
Hydroxyproline	0	2.01	10.8
Proline	23.71	16.35	12.4
Glycine	29.56	29.39	33.3
Aspartic acid	4.91	5.99	4.2
Glutamic acid	8.45	8.07	7.3
Serine	3.02	4.47	3.4
Histidine	0	0.53	0.3
Arginine	4.84	4.53	5.0
Threonine	1.56	2.00	1.6
Alanine	13.38	13.08	11.5
Tyrosine	0.11	0.66	0.1
Valine	2.17	3.09	2.1
Methionine	0.49	0.76	0.7
Isoleucine	0.70	1.35	0.6
Leucine	2.20	2.71	1.9
Phenylalanine	1.32	1.45	1.2
Lysine	3.61	3.57	2.6

* Values based on the literature (*57*)

Glycosylation and Phosphorylation

The amino acid analysis showed 2% hydroxyproline in the corn-derived 44 kDa CIα1. Hydroxylated proline residues represent potential sites for glycosylation. However, neither corn nor *Pichia*-derived 44 kDa CIα1s showed any bands (Figure 6) indicating glycosylation. This glyscosylation staining method can detect the presence of carbohydrate groups in as little as 0.625 ng avidin and 0.16 µg horseradish peroxidase. The amounts of corn and *Pichia*-derived 44 kDa CIα1 loaded to the gel were estimated to be about 2.5 µg, which was about 4000 times that of the detectable avidin and 16 times the detectable horseradish peroxidase.

The Mascot computer algorithm used to process the MS spectrum data allowed us to search for any potential phosphorylated amino acid residues. None were identified in the corn-derived 44 kDa CIα1.

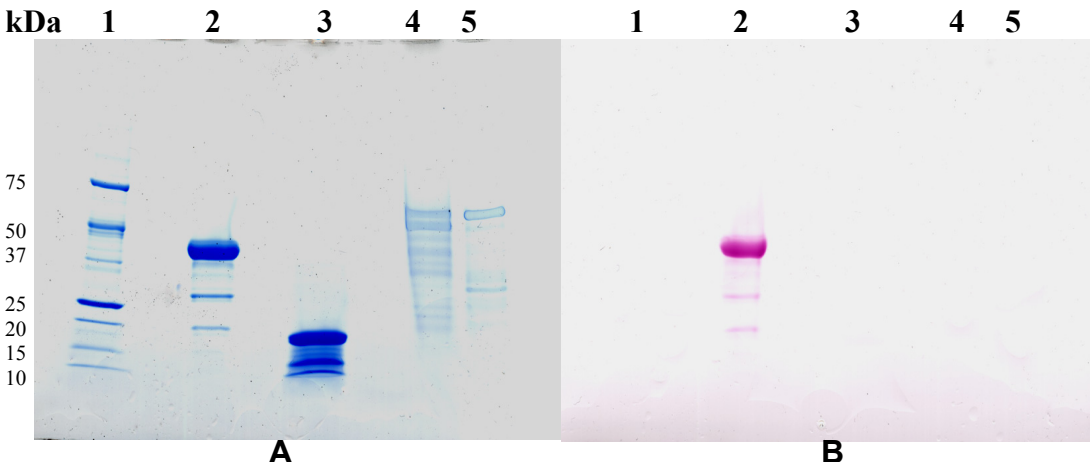

Figure 6. SDS-PAGE gels stained by Coomassie Brilliant Blue (A) or Pierce GelCode glycoprotein staining kit (B). Lane 1, MW markers; Lane 2, HRP (positive control of glycoprotein, 40 kDa); Lane 3, soybean trypsin inhibitor (negative control, 20 kDa); Lane 4, corn grain-derived 44 kDa CIα1 (GFC fraction #2); Lane 5, *Pichia*-derived 44 kDa CIα1.

Conclusions

Protein extraction with 0.1M phosphoric acid buffer at low pH efficiently recovered the corn-derived 44 kDa CIα1 while minimizing the co-extraction of the contaminating host proteins. The combination of CEC and GFC was effective in purifying the corn grain-derived 44 kDa CIα1, present at low concentration in corn extract, at lab scale for biochemical characterization. Data from immunorecognition, MW measurement, internal peptide sequence analysis, N-terminal sequencing, and glycoprotein detection provided reliable evidence that the corn grain-derived 44 kDa CIα1 was faithfully expressed from the recombinant DNA construct transformed into the corn plants with proper cleavage of the signal sequence, a small degree of proline hydroxylation, and no further detectable post-translational modification. This provides first indication of an endogenous corn enzyme able to prolyl-hydroxylate collagenous proteins, a step important to the proteins functional properties.

Acknowledgements

This work was sponsored by the USDA CREES Grant #2003-34496-14026. The authors would like to thank Louisa Tabatabai from Iowa State University and Sheri Almeda from FibroGen for providing technical assistance, the Center for Crops Utilization Research at Iowa State University for use of seed processing equipment, and John Howard from California Polytechnic State University, San Luis Obispo, CA (formerly of ProdiGene) for providing the description of the vector constructs and plant transformation carried out at ProdiGene.

References

(1) Mintert, J. *USDA acreage report surprises trade.* *http://www.lmic.info/memberspublic/InTheCattleMarket/Archives/2007/cattlemarkets0703.pdf;* July 2007

(2) Baker, A.; Zahniser, S. *Ethanol reshapes the corn market.* *http://www.ers.usda.gov/AmberWaves/April06/Features/Ethanol.htm.;* , April, 2006.

(3) USDA. *Adoption of genetically engineered crops in the U.S.* *http://www.ers.usda.gov/Data/BiotechCrops/;* July 2007.

(4) Baez, J.; Olsen, D.; Polarek, J. W. Recombinant microbial systems for the production of human collagen and gelatin. *Appl. Microbiol. Biotechnol.* **2005,** *69*, 245-252.

(5) Olsen, D.; Yang, C. L.; Bodo, M.; Chang, R.; Leigh, S.; Baez, J.; Carmichael, D.; Perala, M.; Hamalainen, E. R.; Jarvinen, M.; Polarek, J. Recombinant collagen and gelatin for drug delivery. *Adv. Drug Delivery Rev.* **2003,** *55*, 1547-1567.

(6) Olsen, D.; Jiang, J.; Chang, R.; Duffy, R.; Sakaguchi, M.; Leigh, S.; Lundgard, R.; Ju, J.; Buschman, F.; Truong-Le, V.; Pham, B.; Polarek, J. W. Expression and characterization of a low molecular weight recombinant human gelatin: development of a substitute for animal-derived gelatin with superior features. *Protein Expr. Purif.* **2005,** *40*, 346-357.

(7) Fertala, A.; Sieron, A. L.; Ganguly, A.; Li, S. W.; Alakokko, L.; Anumula, K. R.; Prockop, D. J. Synthesis of recombinant human procollagen-Ii in a stably transfected tumor-cell line (Ht1080). *Biochem. J.* **1994,** *298*, 31-37.

(8) Fichard, A.; Tillet, E.; Delacoux, F.; Ruggiero, F. Human recombinant alpha 1(V) collagen chain - Homotrimeric assembly and subsequent processing. *J. Biol. Chem.* **1997,** *272*, 30083-30087.

(9) Geddis, A. E.; Prockop, D. J. Expression of human Col1a1 gene in stably transfected Ht-1080 cells - the production of a thermostable homotrimer of type-1 collagen in a recombinant system. *Matrix* **1993,** *13*, 399-405.

(10) Myllyharju, J.; Lamberg, A.; Notbohm, H.; Fietzek, P. P.; Pihlajaniemi, T.; Kivirikko, K. I. Expression of wild-type and modified pro alpha chains of human type I procollagen in insect cells leads to the formation of stable [alpha 1(I)](2)alpha 2(I) collagen heterotrimers and [alpha 1(I)](3) homotrimers but not [alpha 2(I)](3) homotrimers. *J. Biol. Chem.* **1997,** *272*, 21824-21830.

(11) Nokelainen, M.; Lamberg, A.; Helaakoski, T.; Myllyharju, J.; Pihlajaniemi, T.; Kivirikko, K. I. Characterization of human type I, II and III collagens expressed in insect cells. *Matrix Biol.* **1996,** *15*, 194-195.

(12) Cappello, J. The biological production of protein polymers and their Use. *Trends Biotechnol.* **1990,** *8*, 309-311.

(13) Goldberg, I.; Salerno, A. J.; Patterson, T.; Williams, J. I. Cloning and expression of a collagen-analog-encoding synthetic gene in Escherichia-coli. *Gene* **1989,** *80*, 305-314.

(14) Kajino, T.; Takahashi, H.; Hirai, M.; Yamada, Y. Efficient production of artificially designed gelatins with a *Bacillus brevis* system. *Appl. Environ. Microbiol.* **2000,** *66*, 304-309.

(15) Tomita, M.; Munetsuna, H.; Sato, T.; Adachi, T.; Hino, R.; Hayashi, M.; Shimizu, K.; Nakamura, N.; Tamura, T.; Yoshizato, K. Transgenic silkworms produce recombinant human type III procollagen in cocoons. *Nat. Biotechnol.* **2003,** *21*, 52-56.

(16) John, D. C. A.; Watson, R.; Kind, A. J.; Scott, A. R.; Kadler, K. E.; Bulleid, N. J. Expression of an engineered form of recombinant procollagen in mouse milk. *Nat. Biotechnol.* **1999,** *17*, 385-389.

(17) Toman, P. D.; Pieper, F.; Sakai, N.; Karatzas, C.; Platenburg, E.; de Wit, I.; Samuel, C.; Dekker, A.; Daniels, G. A.; Berg, R. A.; Platenburg, G. J. Production of recombinant human type I procollagen homotrimer in the mammary gland of transgenic mice. *Transgenic Res.* **1999,** *8*, 415-427.

(18) Vaughan, P. R.; Galanis, M.; Richards, K. M.; Tebb, T. A.; Ramshaw, J. A. M.; Werkmeister, J. A. Production of recombinant hydroxylated human type III collagen fragment in *Saccharomyces cerevisiae*. *DNA Cell Biol.* **1998,** *17*, 511-518.

(19) Olsen, D. R.; Leigh, S. D.; Chang, R.; McMullin, H.; Ong, W.; Tai, E.; Chisholm, G.; Birk, D. E.; Berg, R. A.; Hitzeman, R. A.; Toman, P. D. Production of human type I collagen in yeast reveals unexpected new insights into the molecular assembly of collagen trimers. *J. Biol. Chem.* **2001,** *276*, 24038-24043.

(20) Geerlings, T. H.; de Boer, A. L.; Lunenborg, M. G. J.; Veenhuis, M.; van der Klei, I. J. A novel platform for the production of nonhydroxylated gelatins based on the methylotrophic yeast *Hansenula polymorpha*. *FEMS Yeast Res.* **2007,** *7*, 1188-1196.

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

(22) Gelse, K.; Poschl, E.; Aigner, T. Collagens - structure, function, and biosynthesis. *Adv. Drug Delivery Rev.* **2003,** *55*, 1531-1546.

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

(24) Ruggiero, F.; Exposito, J. Y.; Bournat, P.; Gruber, V.; Perret, S.; Comte, J.; Olagnier, B.; Garrone, R.; Theisen, M. Triple helix assembly and processing of human collagen produced in transgenic tobacco plants. *FEBS Lett.* **2000,** *469*, 132-136.

(25) Dodge, C.; Baez, J.; Sudarshana, M.; McDonald, K. In *Production of recombinant human gelatin in rice cell cultures*, AAPS Annual Meeting, San Antonio, TX. W5024, 2006.

(26) Holkeri, H.; Wahlstrom, E.; Ritala, A.; Baez, J.; Makinen, K.; Nuutila, A. M. Production of non-hydroxylated recombinant collagen in barley. *J. Biotechnol.* **2005,** *118: S84 (Abstract).*

(27) Ritala, A.; Eskelin, K.; Holkeri, H.; Wahlström, E.; Baez, J.; Mäkinen, K.; Nuutila, A., Development of a model system for the production of a recombinant industrial protein in transgenic barley grain. In *Annual Phytochemical Society of Europe* 2007.

(28) Zhang, C.; Baez, J.; Glatz, C. E., Purification and characterization of a recombinant gelatin expressed in transgenic maize. In *AIChE Annual Meeting*, San Francisco, CA. 425i (Abstract), 2006.

(29) Azzoni, A. R.; Kusnadi, A. R.; Miranda, E. A.; Nikolov, Z. L. Recombinant aprotinin produced transgenic corn seed: Extraction and purification studies. *Biotechnol. Bioeng.* **2002,** *80*, 268-276.

(30) Bailey, M. R.; Woodard, S. L.; Callaway, E.; Beifuss, K.; Magallanes-Lundback, M.; Lane, J. R.; Horn, M. E.; Mallubhotla, H.; Delaney, D. D.; Ward, M.; Van Gastel, F.; Howard, J. A.; Hood, E. E. Improved recovery of active recombinant laccase from maize seed. *Appl. Microbiol. Biotechnol.* **2004,** *63*, 390-397.

(31) Hood, E. E.; Bailey, M. R.; Beifuss, K.; Magallanes-Lundback, M.; Horn, M. E.; Callaway, E.; Drees, C.; Delaney, D. E.; Clough, R.; Howard, J. A. Criteria for high-level expression of a fungal laccase gene in transgenic maize. *Plant Biotechnol. J.* **2003,** *1*, 129-140.

(32) Hood, E. E.; Kusnadi, A.; Nikolov, Z.; Howard, J. A. Molecular farming of industrial proteins from transgenic maize. *Adv. Exp. Med. Biol.* **1999,** *464*, 127-147.

(33) Howard, J. A.; Hood, E. Commercial production of proteases in plants. US Patent 6,087,558, July 11, 2000.

(34) Kusnadi, A. R.; Hood, E. E.; Witcher, D. R.; Howard, J. A.; Nikolov, Z. L. Production and purification of two recombinant proteins from transgenic corn. *Biotechnol. Prog.* **1998,** *14*, 149-155.

(35) Woodard, S. L.; Mayor, J. M.; Bailey, M. R.; Barker, D. K.; Love, R. T.; Lane, J. R.; Delaney, D. E.; McComas-Wagner, J. M.; Mallubhotla, H. D.; Hood, E. E.; Dangott, L. J.; Tichy, S. E.; Howard, J. A. Maize (*Zea mays*)-derived bovine trypsin:

characterization of the first large-scale, commercial protein product from transgenic plants. *Biotechnol. Appl. Biochem.* **2003,** *38*, 123-130.

(36) Lamphear, B. J.; Barker, D. K.; Brooks, C. A.; Delaney, D. E.; Lane, J. R.; Beifuss, K.; Love, R.; Thompson, K.; Mayor, J.; Clough, R.; Harkey, R.; Poage, M.; Drees, C.; Horn, M. E.; Streatfield, S. J.; Nikolov, Z.; Woodard, S. L.; Hood, E. E.; Jilka, J. M.; Howard, J. A. Expression of the sweet protein brazzein in maize for production of a new commercial sweetener. *Plant Biotechnol. J.* **2005,** *3*, 103-114.

(37) Zhong, Q. X.; Xu, L.; Zhang, C.; Glatz, C. E. Purification of recombinant aprotinin from transgenic corn germ fraction using ion exchange and hydrophobic interaction chromatography. *Appl. Microbiol. Biotechnol.* **2007,** *76*, 607-613.

(38) Perret, S.; Merle, C.; Bernocco, S.; Berland, P.; Garrone, R.; Hulmes, D. J. S.; Theisen, M.; Ruggiero, F. Unhydroxylated triple helical collagen I produced in transgenic plants provides new clues on the role of hydroxyproline in collagen folding and fibril formation. *J. Biol. Chem.* **2001,** *276*, 43693-43698.

(39) Gu, Z. R.; Glatz, C. E. A method for three-dimensional protein characterization and its application to a complex plant (Corn) extract. *Biotechnol. Bioeng.* **2007,** *97*, 1158-1169.

(40) Karnoup, A. S.; Turkelson, V.; Anderson, W. H. K. O-Linked glycosylation in maize-expressed human IgA1. *Glycobiology* **2005,** *15*, 965-981.

(41) Belanger, F. C.; Kriz, A. L. Molecular-basis for allelic polymorphism of the maize globulin-1 gene. *Genetics* **1991,** *129*, 863-872.

(42) An, G. H.; Mitra, A.; Choi, H. K.; Costa, M. A.; An, K. S.; Thornburg, R. W.; Ryan, C. A. Functional-analysis of the 3' control region of the potato wound-inducible proteinase inhibitor-Ii gene. *Plant Cell* **1989,** *1*, 115-122.

(43) Anzai, H.; Yoneyama, K.; Yamaguchi, I. Transgenic tobacco resistant to a bacterial disease by the detoxification of a pathogenic toxin. *Mol. Gen. Genet.* **1989,** *219*, 492-494.

(44) Gordonkamm, W. J.; Spencer, T. M.; Mangano, M. L.; Adams, T. R.; Daines, R. J.; Start, W. G.; Obrien, J. V.; Chambers, S. A.; Adams, W. R.; Willetts, N. G.; Rice, T. B.; Mackey, C. J.; Krueger, R. W.; Kausch, A. P.; Lemaux, P. G. Transformation of maize cells and regeneration of fertile transgenic plants. *Plant Cell* **1990,** *2*, 603-618.

(45) Uchimiya, H.; Iwata, M.; Nojiri, C.; Samarajeewa, P. K.; Takamatsu, S.; Ooba, S.; Anzai, H.; Christensen, A. H.; Quail, P. H.; Toki, S. Bialaphos treatment of transgenic rice plants expressing a bar gene prevents infection by the sheath blight pathogen (Rhizoctonia-Solani). *Bio-Technol.* **1993,** *11*, 835-836.

(46) Hiei, Y.; Komari, T. Method for transforming monocotyledons US Patent 5,591,616, May 3, 1994.

(47) Armstrong, C. L.; Green, C. E. Establishment and maintenance of friable, embryogenic maize callus and the involvement of L-proline. *Planta* **1985,** *164*, 207-214.

(48) Vignaux, N.; Octaviani, D.; Fox, S.; Johnson, L. A., Efficiencies of different types of dry mills in recovering a fraction rich in recombinant protein expressed in endosperm. In *Annual Meeting of American Association of Cereal Chemists and the Tortilla Industry Association*, AACC/TIA Annual Meeting Program Book; American Association of Cereal Chemists. St. Paul, MN, 2005; Abstract 305, p 141: San Diego, CA, 2005.

(49) Werten, M. W. T.; Wisselink, W. H.; van den Bosch, T. J. J.; de Bruin, E. C.; de Wolf, F. A. Secreted production of a custom-designed, highly hydrophilic gelatin in *Pichia pastoris*. *Protein Eng.* **2001,** *14*, 447-454.

(50) Werten, M. W. T.; Van den Bosch, T. J.; Wind, R. D.; Mooibroek, H.; De Wolf, F. A. High-yield secretion of recombinant gelatins by *Pichia pastoris*. *Yeast* **1999,** *15*, 1087-1096.

(51) Gao, Y.; Schafer, B. W.; Collins, R. A.; Herman, R. A.; Xu, X. P.; Gilbert, J. R.; Ni, W. T.; Langer, V. L.; Tagliani, L. A. Characterization of Cry34Ab1 and Cry35Ab1 insecticidal crystal proteins expressed in transgenic corn plants and *Pseudomonas fluorescens*. *J. Agric. Food Chem.* **2004,** *52*, 8057-8065.

(52) Adams, W. R.; Huang, S. S.; Kriz, A. L.; Luethy, M. H. Matrix-assisted laser desorption ionization time-of-flight mass spectrometry analysis of zeins in mature maize kernels. *J. Agric. Food Chem.* **2004,** *52*, 1842-1849.

(53) Liebler, D. C., *Introduction to proteomics: tools for the new biology*. Humana Press. p167-184. Totowa, NJ 2002.

(54) Gatlin, C. L.; Eng, J. K.; Cross, S. T.; Detter, J. C.; Yates, J. R. Automated identification of amino acid sequence variations in proteins by HPLC/microspray tandem mass spectrometry. *Anal. Chem.* **2000,** *72*, 757-763.

(55) Johnson, R. S.; Davis, M. T.; Taylor, J. A.; Patterson, S. D. Informatics for protein identification by mass spectrometry. *Methods* **2005,** *35*, 223-236.

(56) de Bruin, E. C.; de Wolf, F. A.; Laane, N. C. M. Expression and secretion of human alpha 1(I) procollagen fragment by *Hansenula polymorpha* as compared to *Pichia pastoris*. *Enzyme. Microb. Technol.* **2000,** *26*, 640-644.

(57) Miller, E. J.; Gay, S. Collagen - an Overview. *Methods Enzymol.* **1982,** *82*, 3-32.

CHAPTER 3. PURIFICATION AND CHARACTERIZATION OF RECOMBINANT COLLAGEN TYPE I alpha 1 FROM TRANSGENIC CORN SEED

A paper to be submitted to Biotechnology Progress

Cheng Zhang[1], Julio Baez[2], and Charles E. Glatz[3]

Abstract

The use of collagen and gelatin produced from animal sources poses quality and safety concerns related to compositional heterogeneity, possible allergic reactions and contamination by animal-derived components. In the present paper we discuss the use of a transgenic corn expression system for the production of a 100 kDa human collagen type I alpha 1 (CIα1) chain that could be developed as a compositionally homogeneous and safe substitute for animal-derived gelatin. While other recombinant proteins have been expressed in corn, producing collagen raises untested post-translational challenges for this host. In addition to obtaining the correct primary sequence, signal sequence processing, proline hydroxylation, and triple helix structure formation were all desired. The expression of CIα1 was directed to the seed by an embryo-specific maize globulin-1 promoter. The corn-derived CIα1 was extracted from ground grain using a low pH buffer (0.1M phosphoric acid, 0.15M NaCl, pH 1.8) to minimize the extraction of endogenous corn contaminating proteins while allowing the efficient recovery of CIα1. The accumulation level of CIα1 was determined to

[1] Primary researcher and author, responsible for purification process development and protein characterization
[2] Co-author, supplying information and materials on collagen/gelatin assays and chemistry
[3] Author for correspondence

be 0.04% of extractable proteins at low pH, creating a challenge for obtaining purified protein for characterization. Available anti-collagen polyclonal antibodies proved ineffective for affinity capture. Thus, purification was carried out by ultrafiltration/diafiltration followed by four chromatographic steps resulting in CIα1 with close to 100% purity and 16% recovery yield. Compared with related animal-derived collagens, the purified recombinant CIα1 had a low level of prolyl hydroxylation (~1.23%). The CIα1 chains folded into a triple-helical structure with a melting temperature around 26^{o}C. The 29 amino acid foldon fused to the C-terminus to initiate triple helix formation was not cleaved from the CIα1chains, but could be removed by pepsin treatment following recovery. The amino acid composition and immunoreactivity of the purified CIα1 was as expected and similar to that of an analogous CIα1 produced in *Pichia*. Tandem mass spectrometry confirmed the primary sequence of the corn-derived 100 kDa CIα1 with 46% coverage. Fragments of the CIα1 chains were also observed, possibly caused by plant endogenous proteases.

Key words: recombinant collagen, human type I collagen, gelatin, protein extraction, purification, characterization, chromatography, transgenic corn, maize.

Introduction

Gelatin, derived from collagen, is extensively used in food, photography, and pharmaceutical applications such as vaccine stabilization, resuscitation fluids, and as the structural component of capsules and tablets. Over 50,000 metric tons of gelatins are consumed every year for these pharmaceutical applications, with over 80% of this material

used to manufacture capsules for the oral delivery of drugs and nutraceuticals. These gelatins are produced mostly from porcine and bovine skins and bones by acid or base extraction (Olsen, et al. 2003). Sourcing gelatins from animals create several concerns. Foremost is a concern for potential allergic reactions and the possibility of pathogenic contaminants, such as prions (Olsen, et al. 2005). In addition, products containing animal-derived components are rejected by a growing number of people due to religious or nutritional preferences. In terms of product quality, animal-derived gelatin lacks source traceability, structural homogeneity, and consistent composition due to the variability of tissue origin and animal age. The use of recombinant technology can deliver custom-tailored recombinant gelatin using a defined, traceable source and process resulting in a product of consistent size, structure, and physical properties. Production of properly-assembled recombinant collagens and gelatin would have numerous applications and their production has been demonstrated in many recombinant systems (Olsen, et al. 2003; Baez, et al. 2005).

For the present study we used corn as the expression system. The use of a crop, such as corn, with well defined genetic and agronomic traits as host is attractive for the production of high-volume (thousand metric tons per year) and low-cost (below US$100/kg) industrial proteins particularly if the recombinant protein can be recovered as a co-product in corn-based biorefineries. We have previously shown that a 44 kDa recombinant human collagen I alpha 1 (CIα1) chain fragment expressed in corn had the expected composition (Zhang, et al. 2008). That molecule was a single chain consisting of ca. 50% of the helical domain of CIα1.

Naturally occurring collagens contain a helical domain with a repeating gly-X-Y composition, with proline as the predominant X or Y amino acid, and non-helical domains flanking the helical domain. These include C- and N-terminal propeptides and C- and N-

terminal telopeptides that play important roles in helping form triple helical structures (Gelse et al., 2003). In the present work, the complete helical region of human collagen I alpha 1 chain was cloned in corn with the telopeptides. The normal C-propeptide, which is thought to be involved in initiating the formation of the triple-helical structure was replaced by a 29 amino acid foldon (Pakkanen, et al. 2003). The foldon was reported to be effective in facilitating the assembly of triple-helical structure in human type I and III collagen molecules. Expression without the N-propeptide did not prevent the formation of correctly folded collagens, suggesting it does not play an important role in collagen assembly (Ruggiero, et al. 2000; Baez, et al. 2005).

One critical element for producing functional collagens is that the recombinant production system can elicit the required post-translational modifications on the nascent collagen to produce a protein that can be used as a replacement for native collagen or gelatin. After their synthesis, the native procollagen chains are directed into the lumen of the endoplasmic reticulum (ER) by the signal peptide, which is then cleaved before assembly of the triple-helical collagen. Before assembly within the ER a specific enzyme, prolyl 4-hydroxylase (P4H), catalyzes the regiospecific hydroxylation of prolines. The extent of hydroxylation determines the thermal stability of the mature triple-helical collagens. The triple-helical procollagen containing the telopeptides and propeptides is converted into mature collagen by removing the N- and C-propeptides. The removal of the propeptides is considered an essential step for subsequent fibril formation (Gelse, et al. 2003).

The currently preferred recombinant system for collagen and gelatin production is fermentation using the yeast *Pichia pastoris* (Olsen, et al. 2003; Baez, et al. 2005). In the *P. pastoris* system, the triple-helical procollagens accumulate within the ER and are not

secreted into the culture medium. Cell lysis and proteolytic treatment (using pepsin) are required for triple-helical collagen recovery (Vuorela, et al. 1997; Keizer-Gunnink, et al. 2000; Myllyharju, et al. 2000; Baez, et al. 2005). The triple-helical structures are resistant to pepsin while non-triple-helical regions and most other host-related proteins are digested (Bruckner and Prockop 1981).

Previous recombinant systems used for collagen production cannot properly hydroxylate the prolines, presumably because the host cells lack sufficient endogenous P4H activity. Because the degree of hydroxylation determines melting temperature of the procollagen and the properties of the derived gelatin, the recombinant system used for collagen production had been modified by cloning the mammalian P4H to perform the required hydroxylation. Collagens expressed in *P. pastoris* (Keizer-Gunnink, et al. 2000) and *Saccharomyces cerevisiae*(Olsen, et al. 2001) were not hydroxylated without co-expression of the P4H. Recombinant human CIα1 expressed in tobacco had a very low content (~0.5%) of hydroxylated prolines, though the low hydroxyproline content did not prevent the correct formation of triple-helical structure (Ruggiero et al., 2000). Co-expression of a heterologous P4H increased the hydroxyproline content in tobacco (Merle, et al. 2002) and yeast (Myllyharju, et al. 2000; Nokelainen, et al. 2001; Pakkanen, et al. 2003).

To characterize the corn-derived CIα1, it was necessary to obtain a pure preparation to determine the amino acid composition and formation of triple-helical structure. The low expression level (~3 mg of CIα1/ kg of seeds) resulting from the first field trial conducted using the early generation corn grain and the lack of a suitable affinity adsorbent challenged this purification. In the present study, we report the development of a purification process consisting of a combination of membrane filtration and conventional non-affinity

chromatography. Previous corn seed extraction studies indicated that the use of acidic pH and optimized ionic strength would minimize the extraction of corn host proteins, thus reducing the complexity of the extract for the subsequent purification (Azzoni, et al. 2002; Zhong, et al. 2007). Purification of recombinant collagens and gelatins produced from other hosts used a combination of precipitation with a single chromatographic step such as ion exchange chromatography (IEC)(Olsen, et al. 2001; Perret, et al. 2001). In some cases, gel filtration chromatography (GFC) (Myllyharju, et al. 2000; Nokelainen, et al. 2001) was used instead of IEC. However, these previous studies had higher CIa1 accumulation (e.g. 0.5 g/l). The combination of low concentration of ~3 mg collagen/kg of corn seeds and the low purity (~0.04%) in the corn extract ruled out the use of the precipitation that had been effective in the earlier work. To deal with this low purity and concentration, a purification process composed of multistep chromatography was pursued in this work.

The characterization of the corn-derived CIα1 included immunoreactivity to CIα1-specific antibodies, amino acid composition, N-terminal sequencing, mass spectrometry and peptide mapping, performed to assess the fidelity of translation and of potential post-translational modifications. Results for the corn grain-derived CIα1 were compared with data from the *Pichia*-derived analogous CIα1. In the present work we 1) show the expression of a full-length collagen α1 chain in cereal for the first time, 2) demonstrate the effectiveness of the purification process that we have developed without affinity chromatography to purify an extremely low abundance protein in a complex corn protein extract for the purpose of protein property characterization, and 3) establish the fidelity of expression including whether the expressed CIα1 chains could be correctly processed to mature collagens and whether corn has endogenous proline hydroxylation activity.

Materials and Methods

Materials

Transgenic corn grain containing CIα1 targeted for expression to the germ was provided by former ProdiGene Inc. (College Station, TX). The CIα1 gene codes for a 1086 amino acid (calculated MW and pI of 98.03 kDa (nominally 100 kDa) and 9.2, respectively) protein, including the N and C telopeptides and the foldon peptide at the C terminus. Upon receipt, the grain was stored at 4^{o}C. A CIα1 derived from the yeast *P. pastoris* (termed as *Pichia*-derived CIα1) with the same sequence was provided by FibroGen Inc. to be used a reference material.

Tris base, sodium chloride, sodium acetate, acetic acid, sodium phosphate dibasic, and phosphoric acid were purchased from Fisher Scientific (Itasca, IL). All chemicals were reagent grade.

Vector Construction and Transformation of Corn CIα1

Vector and transformation were performed by former ProdiGene (College Station, TX). DNA encoding the human 100 kDa CIα1chain (without N-propeptide) was used to make the construct. The C-propeptide in the CIα1 gene was replaced by a 29-amino acid peptide foldon that facilitates chain assembly (Pakkanen, et al. 2003). The ProdiGene corn expression vector PGN9004 was used to clone the gene for CIα1 between the Hind III and Not I sites of the vector. The gene (Figure 1a) consisted of the triple-helical domain, N- and C-telopeptides, and the foldon. Expression was regulated by an embryo-specific maize globulin-1 promoter (PGNpr2) (Belanger and Kriz 1991). The pin II transcription terminator from the protease inhibitor II gene from potato (An, et al. 1989) was used (Figure 1b). For

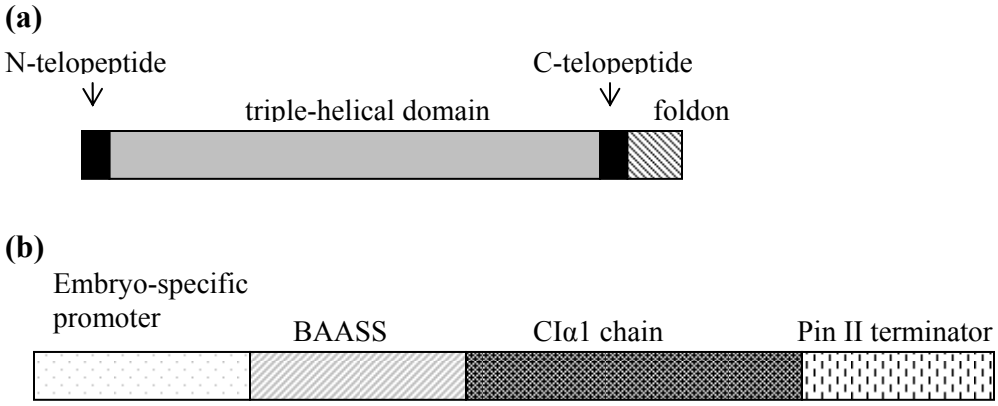

Figure 1. Schematic representations of the recombinant CIα1 chain (a) expressed in corn, and the expression vector (b) for CIα1 (the chain lacks the N-propeptide and the C-propeptide was replaced by a smaller foldon. Domains are not drawn to scale. BAASS=barley alpha amylase signal sequence)

transgenic calli selection the phosphinothricin acetyl transferase gene was included in the vector under the control of the CaMV 35S promoter. This gene confers resistance to the herbicide bialaphos (Anzai, et al. 1989; Gordonkamm, et al. 1990; Uchimiya, et al. 1993). The expression vector also contained the barley alpha amylase signal sequence (BAASS) The vector was introduced into corn immature embryos which were regenerated to recover transgenic plants, as described in the production of the 44 kDa CIα1 chain fragment (Zhang, et al. 2008).

Corn Seed Grinding and Defatting

The CIα1-containing corn kernels were first coarsely milled using a Witt corrugated mill (Witt Corrugating Inc., Wichita, KS), further ground into flour with a household coffee grinder (GE, Fairfield, CT), and then defatted. Defatting was carried out by mixing the ground corn flour with hexane (1:5 w/v) using a magnetic stir bar for 60 min at 0^{o}C. The supernatant was decanted after centrifugation (15 min, 2000*g*, 23^{o}C) and the flour was then

extracted a second time as before. The residual hexane in the flour was removed by air drying at room temperature.

Extraction

Extraction conditions were chosen to extract CIα1 with minimal co-extraction of native host proteins, thus reducing the complexity of the extract for the subsequent chromatographic steps. The defatted corn flour was extracted twice with extraction buffer (0.1M phosphoric acid, 0.15 M sodium chloride, pH 1.8) in a ratio of 1:5 w/v for 1 h at room temperature with stirring. The extracts were clarified by centrifugation (~3000*g*, 10 min, 23^{o}C), then pooled for later concentration and diafiltration by tangential flow filtration (TFF).

Ultrafiltration (UF)/Diafiltration (DF)

Polyethersulfone membranes of 30, 50 and 100 kDa MWCO (PM30, PBQK, PBHK, Millipore Corp, Bedford, MA) were evaluated (Amicon model 8400 ultrafiltration cell, Millipore Corp. Bedford, MA) with the MWCO resulting in highest CIα1 recovery and highest flux used for the preparative procedure. UF/DF was carried out in TFF mode (Pellicon 2 Mini Ultrafiltration Module; Biomax-50 polyethersulfone membranes; 0.1 m^2 area) for concentrating the extract and buffer exchange. The filtration was carried out at constant cross flow rate to obtain ca. eight-fold concentration. The diafiltration was carried with eight diavolumes in a discontinuous mode by repeatedly diluting the sample with an equal volume of dialysis buffer (20 mM sodium phosphate, pH 7.0) followed by two-fold concentration.

Chromatography Development

For cation-exchange chromatography (CEC) three buffers were screened: 50 mM sodium acetate pH 4.8, 20 mM sodium phosphate, pH 7.0, and 50 mM HTris-Cl. The elution was carried out over 15 column volumes (CV) with 0 to 0.5 M sodium chloride gradient in each buffer. The eluting buffer providing highest purity was chosen for the purification.

For the development of a hydrophobic interaction chromatography (HIC) step, a loading buffer salt concentration sought that promoted adsorption but did not cause precipitation of CIα1. Ammonium sulfate was added to the CEC elution pools (to the level of 1.25 M) prior to loading to the HIC column. The buffers that did not cause precipitation at this step were used to screen three HIC resins for best separation performance: Octyl Sepharose™ Fast Flow, Phenyl Sepharose™ Fast Flow (Sigma, St. Louis, MO), and Phenyl Sepharose™ High Performance (GE Healthcare, Uppsala, Sweden). The resin resulting in highest purity of eluted CIα1 was chosen for this step.

***Purification*:**

The purification sequence is shown in Figure 2, with column loading shown in Table 2. Concentrated and dialyzed protein extract was fractionated on a cation-exchange column (XK16/20 column, GE Healthcare, Uppsala, Sweden) packed with 30 ml of SP Sepharose™ Fast Flow resin (Sigma, St. Louis, MO) that had been equilibrated with 20 mM sodium phosphate, pH 7.0 (buffer A). The elution was carried out in stepwise mode from 0 to 0.2 M sodium chloride in buffer A over 3.5 CV at 3 ml/min flow rate. The fraction with highest CIα1 content was dialyzed (Pierce SnakeSkin T Dialysis Tubing, 10K MWCO, Pierce, Rockford, IL) overnight at 4^{o}C against buffer A.

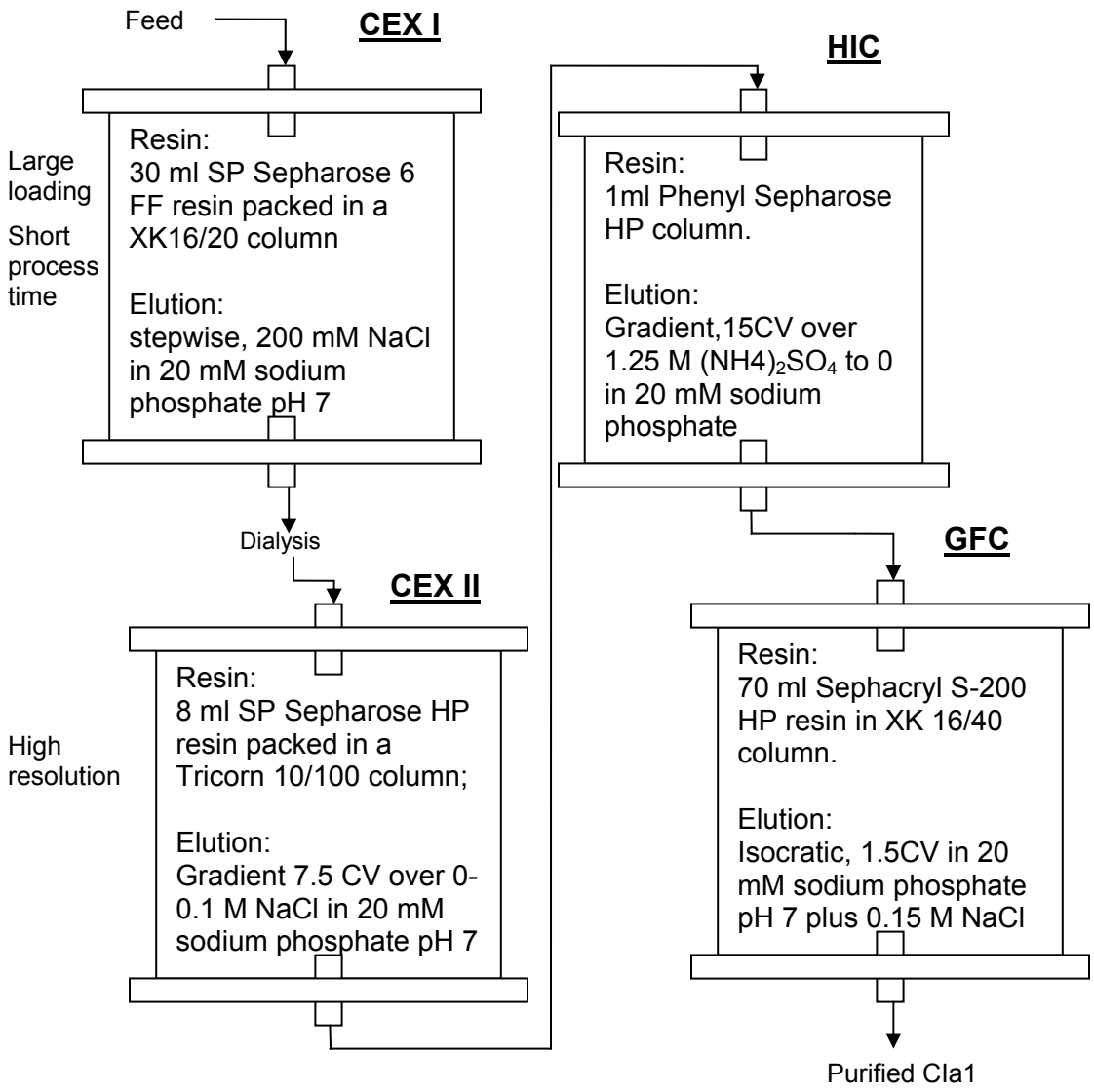

Figure 2. Flow diagram and running conditions for the chromatographic steps of the corn-derived CIα1 purification from concentrated and dialyzed corn extract.

The dialyzed fraction was applied onto a Tricorn 10/100 column (GE Healthcare, Uppsala, Sweden) packed with 8 ml SP Sepharose High Performance resin (GE Healthcare, Uppsala, Sweden). The elution was carried out over 7.5 CV of 0 to 0.1 M sodium chloride gradient in buffer A at 1 ml/min flow rate. Fractions containing CIα1 were pooled and 2.5 M ammonium sulfate in buffer A was added to a final concentration of 1.25 M ammonium sulfate. The solution was loaded to a HIC column packed with 1 ml of Phenyl Sepharose™ High Performance resin (HiTrap Phenyl HP, GE Healthcare, Uppsala, Sweden). The elution was carried out over 15 CV of 1.25 to 0 M ammonium sulfate gradient in buffer A. Fractions containing CIα1 were pooled and concentrated with a centrifugal filter (Amicon Ultra-4, Millipore, Bedford, MA). The concentrated sample (~ 1 ml) was then loaded onto a XK 16/40 column (GE Healthcare, Uppsala, Sweden) packed with 70 ml of Sephacryl S-200HR GFC medium (GE Healthcare, Uppsala, Sweden). Proteins were eluted in buffer A plus 0.15 M sodium chloride at 0.5 ml/min flow rate. The fraction with highest CIα1 content was used for characterization.

Protein Sequencing by Tandem Mass Spectrometry (MS)

The purified corn-derived CIα1 was denatured in 6M guanidine-HCl, 50 mM Tris-HCl (pH 8) and 2mM dithiothreitol (DTT) at 60^{o}C for 45 min followed by protease (sequencing grade trypsin, Promega, Madison, WI) digestion for 3 h at 37^{o}C at an enzyme to substrate ratio of 1:20 (w/w) in 50 mM ammonium bicarbonate buffer, pH 7.8. The digest was diluted 2X with 50% MeOH and 0.5% formic acid and the resultant solution was purified and concentrated with C18 ZipTip (Millipore, Bedford, MA). About 20 µl of sample was loaded to an ESI-quadrupole-TOF tandem mass spectrophotometer (Qstar XL,

Applied Biosystems, Foster City, CA). Full scan mass spectra were acquired and used to determine the precursor ions which were further fragmented to generate MS/MS spectra. Analysis of the MS and tandem MS (MS/MS) spectral data was performance using the Mascot computer algorithm, which allows for peptide amino acid sequence interpretation. A self-defined database of the human CIα1 chain amino acid sequence and the adjoining sequences used in the gene construct was searched. The trypsin digestion, MS spectra collection and amino acid sequence interpretation were carried out in the Proteomics Facility at Iowa State University (ISU).

Protein Assay

Total protein concentration was determined using the Pierce's Coomassie Plus - The Better Bradford™ Assay Kit (Pierce, Rockford, IL), with bovine serum albumin as the standard. The assay was done in triplicate and reported as the average.

Enzyme-Linked Immunosorbent Assay (ELISA)

CIα1 concentrations were determined by a competitive ELISA assay in which heat-denatured *Pichia*-derived 100 kDa CIα1 (concentration:50 ng/ml) was immobilized at 4^{o}C overnight in a 96-well plate (Costar High Binding Cat# 3590, Corning, NY) as a competitor against corn-derived CIα1 for the primary antibody. The primary antibody was a rabbit polyclonal antibody (#CA725, FibroGen, South San Francisco, CA), raised against a *Pichia*-derived 25kDa CIα1, and used at 1:4000 dilution. The 25kDa CIα1 segment was part of the CIα1 and this antibody was reactive with both. Horseradish peroxidase (HRP) –conjugated goat anti-rabbit IgG (H+L) (Invitrogen, Carlsbad, CA) was used as the secondary antibody at

a 1:5000 dilution. Both primary and secondary antibodies were diluted in the assay buffer (100 mM PBS, 0.05% Tween 20, 0.1% BSA, 0.1% Kathon CG/ICP, pH 7.0). A SureBlue™ TMB substrate solution (Kirkegaard & Perry Laboratories, Gaithersburg, MA) was used as the substrate for the HRP conjugate and incubated with it at ambient temperature (23^{o}C) for 30 min. The plate was then read at 450 nm on a microplate reader (EL340, BioTek Instruments, Winooski, VT). *Pichia*-derived CIα1 of known concentration was the reference material for the quantification of the corn grain-derived CIα1.

SDS-PAGE and Western Blotting

SDS-PAGE was carried out according to the Ready Gel® Precast Gels: Application Guide (BioRad, Hercules, CA) in a Mini-PROTEAN® II Electrophoresis Cell (Bio-Rad, Hercules, CA). Proteins samples were diluted in equal volume of sample buffer (83 mM Tris-HCl, pH 6.8, 2.7% (w/v) SDS, 25% glycerol, 0.02% Bromophenol Blue ,7% 2-mercaptoethanol) and heated at 100^{o}C for 5 min before being loaded to a 4-15% Tris-HCl precast gel (Bio-Rad, Hercules, CA). After electrophoresis, the gels were silver stained (Bio-Rad, Hercules, CA).

In Western blotting analysis, the primary antibody used was either the anti-25kDa antibody as described above or an anti-foldon antibody (#CA 74550, FibroGen, South San Francisco, CA), and secondary antibodies used were the same as for the ELISA assay. After electrophoresis, the gel was transferred to a PVDF membrane (Invitrogen, Carlsbad, CA) and an ECL plus Western blotting detection system (Amersham Biosciences, RPN 2132, Buckinghamshire, UK) was used to visualize CIα1 bands. The membrane was exposed to a Fuji Super Rx autoradiographic film (Fisher Scientific, 0444199, Itasca, IL) for about 30 sec

and the film was subsequently developed in a Futura^{TM} 2000K film processor (Fischer Industries, Inc. Geneva, IL).

N-terminal Sequencing and Amino Acid Analysis

The N-terminal sequencing was performed at the ISU Protein Facility. The corn-derived CIα1 fraction from the GFC column was loaded onto the Procise protein sequencer (Applied Biosystems, Foster City, CA). Cycles of Edman degradation were conducted followed by injection onto a RP-HPLC.

For the amino acid analysis (also performed at the ISU Protein Facility), the purified corn-derived CIα1 was loaded to SDS-PAGE and transferred to a PVDF membrane. The excised CIα1 band was subjected to hydrolysis under vacuum (6 N HCl, 150^{o}C, 65 min). The resultant amino acids were then derivatized with phenylisothiocyanate (PITC) by using a PTC derivatizer (Perkin Elmer, Applied Biosystems Model 420A, Foster City, CA), and analyzed by a PTC Amino Acid Analyzer (Perkin Elmer, Applied Biosystems Model 130A, Foster City, CA). A *Pichia*-derived CIα1 (in 10 mM HCl solution instead of a PVDF band) was also submitted and analyzed for comparison. *Trans*-4-hydroxy-L- proline (Sigma H5534), *cis*-4-hydroxy-L-proline (sigma H1637), and *trans*-3-hydroxy-L-proline (Sigma 56244) were used as standards for quantifying the hydroxyproline content. The *cis*-4-proline and *trans*-3-proline standards were run at a later time than the samples, and the retention times were normalized with respect to the trans-4 standard run both times.

Pepsin Resistance Study and Melting Temperature (T_m) Determination

The method to determine T_m is based on the pepsin-digestion resistance of the triple-helix form. Below its T_m, collagen is expected to have a triple-helical structure where pepsin does not have access to its cleavage sites. Comparison of SDS gels before and after digestion reveals whether the sample was digested (i.e. was melted) or not (not melted). Purified corn-derived CIα1(~0.6-0.7 µg) was incubated for 15 min in a Polystat® refrigerated bath (EW-12111-10, Cole Parmer, Vernon Hills, IL) at desired temperatures, and then treated with freshly prepared 150 µg/ml porcine pepsin solution (Sigma P6887, St. Louis, MO) at 2-8ºC for 16-18 h. Reactions were terminated by raising the pH above 6 with 2M Tris base. The digests were analyzed by Western blotting. T_m was determined as the temperature when the CIα1 band disappeared. For comparison, two *Pichia*-derived CIα1s (nonhydroxylated and hydroxylated) were also digested by pepsin, but analyzed with silver staining to determine T_m.

Results and Discussion

Expression of CIα1 in Corn

With the intention to achieve high expression level in corn seeds, an embryo-specific maize globulin-1 promoter and a barley alpha amylase signal sequence (BAASS) were used in this study. The extractable CIα1 in the corn grain was found to be 3 mg/kg grain by ELISA. The same promoter and signal sequence combination had provided higher expression levels of other recombinant proteins in transgenic corn: laccase with accumulation level of 51 mg/kg seed (Hood, et al. 2003; Bailey, et al. 2004), and brazzein of 400 mg/kg seed (Lamphear, et al. 2005). Feasible commercial production would require accumulation above

0.1% (1 g/kg seeds) (Nandi, et al. 2005).

Protein Extraction

Our previous study to purify a 44 kDa CIα1 chain fragment used a 3-h extraction time (Zhang, et al. 2008). Extraction beyond 1 h did not yield any changes in either CIα1 or total extractable protein concentration in the extract. The content of the CIα1 in the extract was very low, about 0.25 μg/ml or 0.04% of the total soluble protein (TSP) in the extract. The TSP extracted at this acidic condition (0.1 M phosphoric acid, 0.15 M NaCl, pH as is (~1.8)) was 0.6 mg/ml or 4.3 mg/ g seed, which is consistent with previous corn protein extraction studies (Azzoni, et al. 2002; Zhong, et al. 2007; Zhang, et al. 2008). The extraction condition maximized the CIα1concentration, as the CIα1 has higher solubility in acid pH than in neutral pH, and minimized the amount of soluble corn protein in the extracts, thus reducing the subsequent purification burden. Given that expression was targeted to the germ, dry-milling to separate a germ-rich fraction would enrich and concentrate CIα1 prior to protein extraction, but this was not pursued in the present work.

Protein Purification

Of the three membranes tested for the initial UF/DF step, the 50kDa MWCO membrane was selected for its higher flux rate (thus shorter processing time) and no loss of CIα1 in the permeate. The TFF, as shown in Table 1, reduced the extract volume from about 2000 ml to about 250 ml (8X volume reduction). In addition, it removed more than 80% of the contaminating proteins, providing a purification factor close to 4. Thus, loading time and resin volume were both reduced for the subsequent chromatographic step. Further, DF

Table 1. Summary of ultrafiltration and diafiltration (UF/DF) of corn extract

	Collagen (µg/ml)	Vol (ml)	Purity (%)	Purification factor	Step yield (%)
Protein Extract	0.25	2000	0.04	—	—
Retentate	0.59	250	0.15	3.8	30-50

provided the required buffer exchange. The yield for the UF/DF step was only 30-50% due to adsorption of the CIα1 on the membrane surface (the CIα1 content was negligible in the permeate). A possible improvement would be to include a washing step or evaluate other membrane chemistry.

Table 2 summarizes the outcomes of each chromatographic step. A cation-exchange resin (sulphopropyl) was chosen to take advantage of the high pI (9.2) of CIα1. Of the three buffers/pHs tested in the CEC step development, the 20 mM sodium phosphate pH 7.0 gave the highest CIα1 purity and comparable yields in the CEC elution pools (data not shown). CEC was carried out in two steps providing step enrichments of 5.7 and 4.3 (Table 2). Two CEC steps were used because SP Sepharose FF resin in stepwise elution mode could accommodate the large sample loadings at high flow rate and the SP Sepharose HP resin in gradient elution mode could achieve high resolution.

For the HIC step development, a small scale solubility test showed that ammonium sulfate concentration as high as 1.25 M did not cause any precipitation of CIα1. Therefore, two concentrations, 1.25 and 0.625 M, were evaluated for adsorption. While 0.625 M, led to CIα1 loss in the flow-through pool, 1.25 M ensured binding on all three HIC resins used. Phenyl Sepharose HP resin was selected for the final HIC step as it showed superior performance in both yield and step purification factors (Table 3). After the CEC steps, the

Table 2. Summary of chromatographic steps of the purification process of corn-derived CIα1

Step	Volume (ml)	CIα1 (μg)	Protein (μg)	Purity %*	Step Purification Factor**	Step Yield (%)
Feed	250	149	101,125	0.15	—	—
CEC I	66	119	13,794	0.86	5.7	80
CEC II	16	76	2,049	3.71	4.3	76
HIC	9	62	177	35.0	9.4	87
SEC	5	68	66	103	2.9	100
				Overall	668	53

* Purity is based on the amount of 100 kDa CIα1 as determined by ELISA and total protein as determined by Coomassie blue protein assay using BSA as a reference standard.

** Step purification factor is determined by the purities before and after the unit operation.

Table 3. HIC step resin screening result

Salt Concentration (m)	**Type of Resin**	**Step Yield (%)**	**Step Purification Factor**
1.25	Octyl FF	70	2
	Phenyl FF	85	4.5
	Phenyl HP	87	9.4

CIα1in the HIC pool was further enriched 9.4-fold (Table 2).

CIα1 was collected in a single fraction (the chromatogram, not shown, evidenced a double peak) well before the elution of native corn proteins in the GFC step. The longer retention time of the native proteins is consistent with evidence that ca. 60% of corn proteins extracted at acidic pH are 3-4 times smaller than the single CIα1chain (98 kDa) (Gu and Glatz 2007).

The four chromatographic steps, in combination with UF/DF, achieved a combined purification factor of 2500 with an overall yield of 16%. The purification process was also monitored by SDS-PAGE (Figure 3a) and the identity of the purified CIα1was confirmed by Western blotting (Figure 3b). While UF/DF (Lane 3, Figure 3a) removed most of contaminating host proteins, CEC (Lane 4) enriched both contaminating corn proteins and the product protein (CIα1). HIC (Lane 5) and GFC (Lane 6) removed impurities, yielding the purified CIα1.

MW of CIα1

The MW of corn-derived CIα1 estimated from the gel (lane 5, Figure 3(A)) was about 120 kDa, 20% higher than the expected MW of 100 kDa for a single CIα1 chain. Such a discrepancy has been reported for collagens frequently and a variety of explanations based on collagen's lower average amino acid size (Butkowski, et al. 1982), reduced chain flexibility (Manca, et al. 1997), and higher hydrophilicity (Werten, et al. 2001) have been used to explain this common observation. Additionally, two or three faint protein bands appear below the main "120 kDa" band (Lane 5). These could indicate that variable posttranslational cleavages have occurred. Multiple bands did not show up in the corresponding Western blot

(Figure 3 (B)); however, other bands did show immuno-reactivity (Figure 4) when higher sample loadings were used. Apparent molecular weights of these bands ranged from 120 to 70 kDa, consistent with the variable cleavage possibility. The smaller fragments proved, by their susceptibility to pepsin digestion, that they were not able to form stable triple helices (Figure 5, panel B, lane 1).

N-terminal Sequencing

N-terminal sequence analysis revealed a peptide sequence "GXPGPXGAPGP", which was nine amino acids downstream of the N-terminus of the triple-helical domain, providing more evidence that unintended proteolytic cleavage occurred. Several recombinant proteins expressed in corn with the same targeting via ER as used here have shown evidence of cleavage in the structural domain, including a 44 kDa CIα1chain fragment (Azzoni, et al. 2002; Zhong, et al. 2007; Zhang, et al. 2008), E1 endocellulase and cellobiohydrolase I (Hood, et al. 2007). Since the content of the corn-derived CIα1 in the extract was very low, the fragments were not detectable until they were purified and concentrated. Therefore, it is uncertain whether the fragments were generated during the growth of the corn seeds, grain storage or the extraction and purification process. If the fragments were generated during the lengthy multistep purification process by concomitantly released plant proteases from extraction, inhibitors could be added during the protein extraction.

Assembly of Triple-Helical Structure

The continuing presence of intact CIα1 chains after pepsin digestion (Figure 5) is evidence of proper triple helix assembly and, in turn, the effectiveness of the foldon replacing

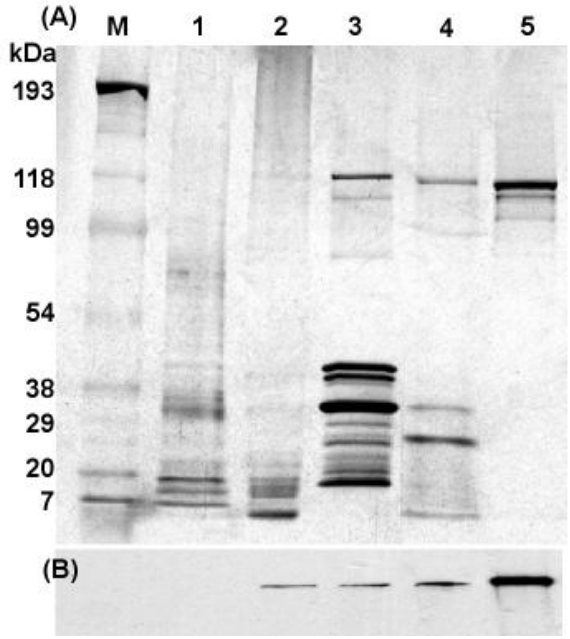

Figure 3. SDS-PAGE of fractions thoughout the purification process visualized with silver staining (A), and western blotting with an anti-25 kDa **CIα1** antibody (B). Lane M, protein standard marker; lane 1, extract; lane 2, retentate of TFF; lane 3, CEC II pooled fractions; lane 4, HIC pooled fractions; lane 5, GFC pooled fractions. The bands in B are at the position corresponding to a molecular weight of 120 kDa.

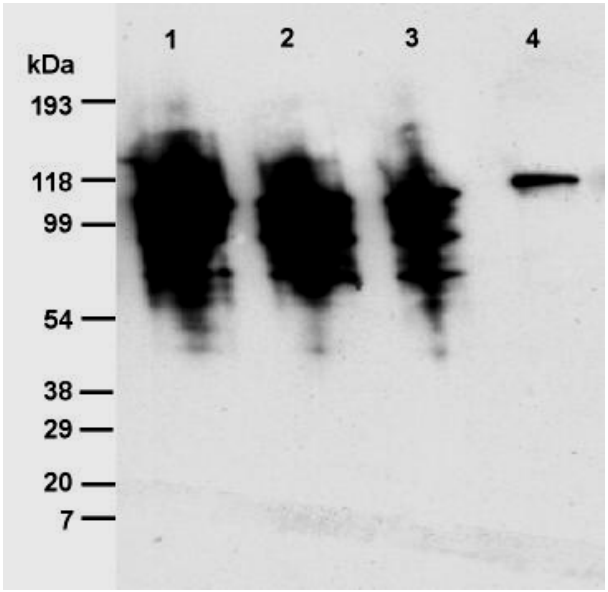

Figure 4. Western blot of purified CIα1 for different sample loadings. Lane 1: 0.5µg , Lane 2: 0.3µg, Lane 3: 0.2µg, and Lane 4: 0.03µg). Higher loadings reveal the presence of some smaller chains.

the native C-propeptide. Also shown is that the *Pichia*-derived nonhydroxylated and hydroxylated CIα1s containing foldon are pepsin-resistant (Pakkanen, et al. 2003; Pakkanen, et al. 2006). Without either the foldon or the C-propeptide, CIα1 produced from *P. pastoris* was reported to form unstable triple-helical structure (Pakkanen, et al. 2003). The presence of the foldon in the corn-derived CIα1 was confirmed by recognition with anti-foldon antibody in Western blotting (Figure 6, lane 1). The foldon was removed by pepsin treatment (Figure 6, lane 2). The presence of the foldon indicated that C-terminal processing (foldon cleavage) did not occur in the seed or during processing (extraction and purification). The same retention of the foldon occurred in *P. pastoris* (Pakkanen, et al. 2003). Although effective in triple helix assembly, neither the C-propeptide nor the foldon was processed (removed) by the *P. pastoris,* and subsequent pepsin treatment was required to convert the procollagens to mature collagens (Myllyharju, et al. 2000; Nokelainen, et al. 2001; Pakkanen, et al. 2003; Baez, et al. 2005). Interestingly, CIα1 expressed in tobacco was processed to mature collagen simultaneously by concomitantly released endogenous tobacco proteases, thus eliminating the need for pepsin to cleave the N- and C-propeptides (Ruggiero, et al. 2000). The differences in proteolytic action between corn and tobacco could be due to either the differences in the extraction conditions or in their endogenous proteases.

Further comparison of the CIα1 chains without pepsin treatment (Figure 5, panel a) to those with pepsin treatment (Figure 5, panel b) revealed that all *Pichia*-derived hydroxylated CIα1chains were resistant to pepsin digestion, as the protein band intensity remained the same before and after pepsin treatment. In contrast, *Pichia*-derived non-hydroxylated CIα1 chains and corn-derived chains were liable to pepsin digestion to some degree. The corn CIα1 chains were more susceptible to pepsin degradation as distinctly faint bands were

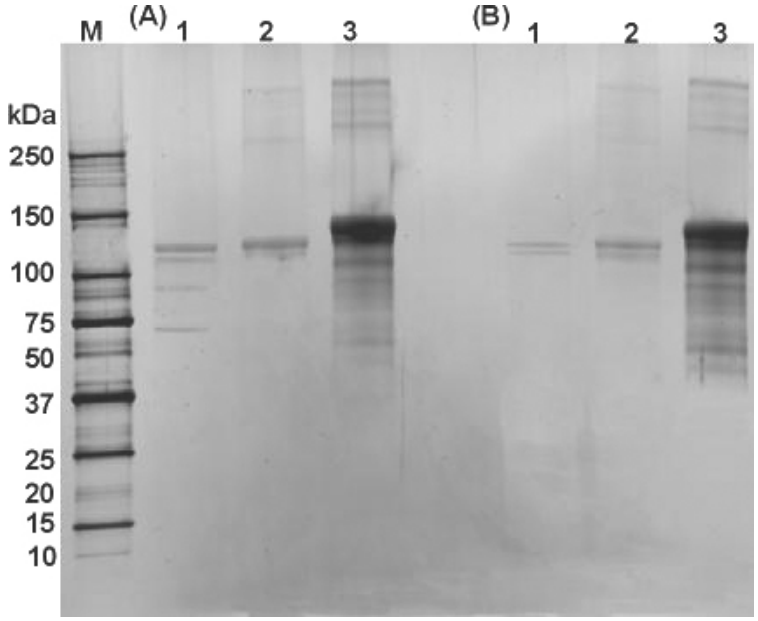

Figure 5. Analysis of the assembly of CIα1 chains into triple-helical structures. Corn-derived CIα1(lanes 1), Pichia-derived non hydroxylated CIα1(lanes 2), and hydroxylated CIα1chains (lanes 3) were analyzed without pepsin treatment (A) and after pepsin digestion for 17 h at 4 °C (B) by 4-15% SDS-PAGE under reducing conditions followed by silver staining. M represents a molecular weight marker.

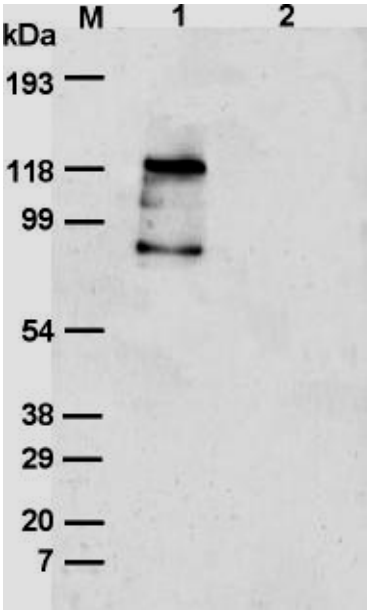

Figure 6. Western blot of corn-derived CIα1 with anti-foldon antibody before (lane 1) and after pepsin treatment (lane 2). (The molecular weight scale (shown as lane M) was derived from the original SDS-PAGE gel MW standards lane)

observed after pepsin treatment (Figure 5, panel b, lane 1), which indicated that the only a fraction of CIα1 chains produced by the transgenic corn assembled into the triple-helix, with the rest either forming deficient triple-helical structure or remaining as single chains. Whether the low hydroxyproline content (as will be shown below) hinders the correct triple-helical structure formation, or the plant endogenous proteases have already cleaved some of the nascent CIα1 chains before they have a chance to assemble into the correct triple-helical structure is unknown.

Amino Acid Composition, Proline Hydroxylation and Thermal Stability

Amino acid analysis (Table 4) showed that the corn-derived CIα1 is similar to its *Pichia*-derived hydroxylated CIα1 and a human CIα1 for all the amino acids measured except for proline and hydroxyproline. The high levels of proline (ca. 20%) and glycine (ca. 30%) are uniquely characteristic of collagens. The hydroxyproline content for corn-derived CIα1 is 1.23% of the total amino acids, corresponding to about 10% of the hydroxylated proline in native type I collagens. This result is slightly higher than that observed in tobacco-derived CIα1(Ruggiero, et al. 2000), suggesting that a naturally occurring corn P4H is able to modify proline but only to a limited extent. Since hydroxylation of proline residues in collagen is an important post-translational modification and known to be critical for the thermal stability of the triple-helical domain, co-expression of heterologous proline hydroxylase with CIα1 would be necessary in transgenic plants for higher hydroxyproline content and thermal stability as reported previously in tobacco cell culture (Olsen, et al. 2003). All hydroxyproline was *trans*-4-hydroxyproline as neither *cis*-4-hydroxyproline nor *trans*-3-hydroxyproline was detected.

Table 4: Amino acid composition of CIα1 from several sources

Amino acid	Corn CIα1 (%)	*Pichia* OH-CIα1 (%)	Human type I homotrimer * (%)
Hydroxyproline	1.23	11.54	10.8
Proline	18.50	11.22	12.4
Glycine	27.82	29.6	33.3
Aspartic acid	5.04	5.58	4.2
Glutamic acid	7.71	7.97	7.3
Serine	4.43	3.75	3.4
Histidine	0.75	0.32	0.3
Arginine	5.14	5.40	5.0
Threonine	2.20	2.17	1.6
Alanine	12.45	10.93	11.5
Tyrosine	0.91	0.47	0.1
Valine	3.35	1.99	2.1
Methionine	0.74	0.72	0.7
Isoleucine	1.53	0.89	0.6
Leucine	2.52	2.29	1.9
Phenylalanine	1.55	1.54	1.2
Lysine	4.12	3.62	2.6

* Values based on the literature (Miller and Gay 1982)

The melting temperature (T_m), an index of the thermal stability, for the corn-derived CIα1 is estimated to be 26-27^{o}C based on the pepsin digestion study (Figure 7A). In comparison, *Pichia*-derived counterparts had T_m around 24-25^{o}C for non-prolyl hydroxylated (Figure 7B) and 38-39^{o}C for fully prolyl hydroxylated (Figure 7C). Since the T_m of the collagens has been reported to be directly related to the hydroxyproline content (Ruggiero, et al. 2000; Merle, et al. 2002), the difference in T_m confirms the differences in hydroxyproline content. As evidenced in Table 4, the hydroxyproline contents for corn-derived and *Pichia*-derived CIα1s were 1.23 and 11.22%, respectively. Since the co-expression of P4H succeeded in increasing hydroxyproline content in other hosts, including tobacco cell culture (Myllyharju, et al. 2000; Merle, et al. 2002), the same strategy could be applied in transgenic corn. By controlling the co-expression of P4H with CIα1, the level of proline hydroxylation could be tailored to suit different application.

Internal Peptide Sequencing by Tandem MS

As shown in Figure 8, the confirmed sequence coverage from the MS data was 46%. This value was enough to confirm in the CIα1 identity, as literature uses as little as 10-15% sequence coverage to verify a protein being expressed (Liebler 2002). The MS data showed the existence of the telopeptide, resolved as the first peptide highlighted in red bold in Figure 8. However, the N-terminal analysis data best matched the segment beginning nine-amino acids downstream of the N-terminal telopeptide. The product may be a mix of collagens differing in the point of N-terminal processing as Figures 4 and 5 showed evidence of multiple forms in this preparation. However, N-terminal sequencing using the single excised band of the Figure 5 gel did not provide a matching sequence.

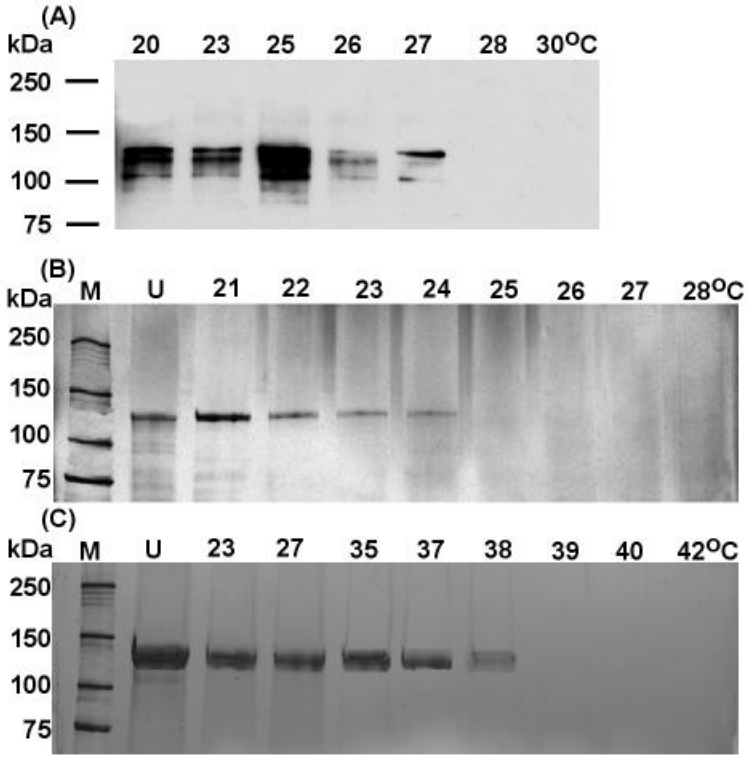

Figure 7. Analysis of the thermal stability of the CIα1 chains at different temperatures. Corn-derived CIα1(A), *Pichia*-derived non hydroxylated CIα1(B), and *Pichia*-derived hydroxylated CIα1chains (C) were heated treated at specified temperature for 15 minutes, digested with pepsin for 17 h at 4°C, and subsequently analyzed by SDS-PAGE with silver staining (B and C) or Western blotting (A). M: protein standard marker, U: not digested.

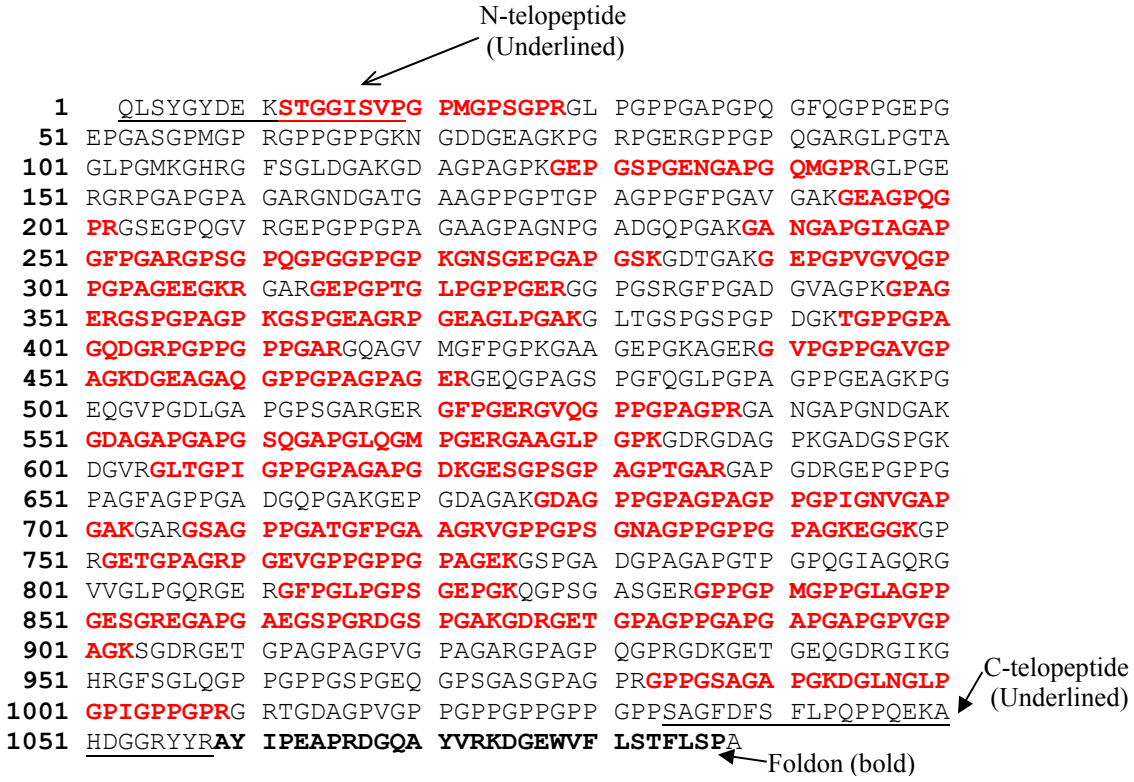

Figure 8. Sequence coverage 46% (peptides in red bold) resolved by tandem MS with trypsin digestion. N- and C-telopeptides and foldon are underlined.

Conclusions

The combination of membrane filtration and four-step chromatographic steps was effective in purifying the corn-derived 100 kDa CIα1 present in low amount in corn extract. Data from immunoreactivity, MW estimation, amino acid composition and internal peptide sequence analysis provided reliable evidence that the corn-derived CIα1 was consistent with the DNA construct transformed into the corn plants. A fraction of the expressed CIα1 formed triple helices resistant to pepsin hydrolysis. The relatively low hydroxyproline content in the corn-derived CIα1 confirmed both some ability of corn to hydroxylate proline, but also the necessity of co-expression of a heterologous P4H if higher content is desired. Higher expression in future generations would simplify the purification process and perhaps reduce the proteolytic cleavage that may have occurred during this lengthy purification process.

Acknowledgements

The authors thank Louisa Tabatabai from Iowa State University (ISU) and Sheri Almeda from FibroGen for providing technical assistance, Siquan Luo from the Proteomics Facility at ISU for performing MS analysis, the Center for Crops Utilization Research at Iowa State University for use of seed processing equipment, and John Howard from California Polytechnic State University, San Luis Obispo, CA (and formerly of ProdiGene) for providing the description of the vector constructs and plant transformation carried out at ProdiGene. This work was supported by the funding from the USDA CREES Grant #2006-34496-17122.

References

An, G. H.; Mitra, A.; Choi, H. K.; Costa, M. A.; An, K. S.; Thornburg, R. W.; Ryan, C. A. Functional-analysis of the 3' control region of the potato wound-inducible proteinase inhibitor-Ii gene. *Plant Cell* **1989,** 1 (1), 115-122.

Anzai, H.; Yoneyama, K.; Yamaguchi, I. Transgenic tobacco resistant to a bacterial disease by the detoxification of a pathogenic toxin. *Mol. Gen. Genet.* **1989,** 219 (3), 492-494.

Azzoni, A. R.; Kusnadi, A. R.; Miranda, E. A.; Nikolov, Z. L. Recombinant aprotinin produced transgenic corn seed: Extraction and purification studies. *Biotechnol. Bioeng.* **2002,** 80 (3), 268-276.

Baez, J.; Olsen, D.; Polarek, J. W. Recombinant microbial systems for the production of human collagen and gelatin. *Appl. Microbiol. Biotechnol.* **2005,** 69 (3), 245-252.

Bailey, M. R.; Woodard, S. L.; Callaway, E.; Beifuss, K.; Magallanes-Lundback, M.; Lane, J. R.; Horn, M. E.; Mallubhotla, H.; Delaney, D. D.; Ward, M.; Van Gastel, F.; Howard, J. A.; Hood, E. E. Improved recovery of active recombinant laccase from maize seed. *Appl. Microbiol. Biotechnol.* **2004,** 63 (4), 390-397.

Belanger, F. C.; Kriz, A. L. Molecular-basis for allelic polymorphism of the maize globulin-1 gene. *Genetics* **1991,** 129 (3), 863-872.

Bruckner, P.; Prockop, D. J. Proteolytic-enzymes as probes for the triple-helical conformation of procollagen. *Anal. Biochem.* **1981,** 110 (2), 360-368.

Butkowski, R. J.; Noelken, M. E.; Hudson, B. G. Estimation of the Size of Collagenous Proteins by Electrophoresis and Gel Chromatography. *Methods Enzymol.* **1982,** 82, 410-423.

Gelse, K.; Poschl, E.; Aigner, T. Collagens - structure, function, and biosynthesis. *Adv. Drug Delivery. Rev.* **2003,** 55 (12), 1531-1546.

Gordonkamm, W. J.; Spencer, T. M.; Mangano, M. L.; Adams, T. R.; Daines, R. J.; Start, W. G.; Obrien, J. V.; Chambers, S. A.; Adams, W. R.; Willetts, N. G.; Rice, T. B.; Mackey, C. J.; Krueger, R. W.; Kausch, A. P.; Lemaux, P. G. Transformation of maize cells and regeneration of fertile transgenic plants. *Plant Cell* **1990,** 2 (7), 603-618.

Gu, Z. R.; Glatz, C. E. A method for three-dimensional protein characterization and its application to a complex plant (Corn) extract. *Biotechnol. Bioeng.* **2007,** 97 (5), 1158-1169.

Hood, E. E.; Bailey, M. R.; Beifuss, K.; Magallanes-Lundback, M.; Horn, M. E.; Callaway, E.; Drees, C.; Delaney, D. E.; Clough, R.; Howard, J. A. Criteria for high-level

expression of a fungal laccase gene in transgenic maize. *Plant Biotechnol. J.* **2003,** 1 (2), 129-140.

Hood, E. E.; Love, R.; Lane, J.; Bray, J.; Clough, R.; Pappu, K.; Drees, C.; Hood, K. R.; Yoon, S.; Ahmad, A.; Howard, J. A. Subcellular targeting is a key condition for high-level accumulation of cellulase protein in transgenic maize seed. *Plant Biotechnol. J.* **2007,** 5 (6), 709-719.

Keizer-Gunnink, I.; Vuorela, A.; Myllyharju, J.; Pihlajaniemi, T.; Kivirikko, K. I.; Veenhuis, M. Accumulation of properly folded human type III procollagen molecules in specific intracellular membranous compartments in the yeast *Pichia pastoris*. *Matrix Biol.* **2000,** 19 (1), 29-36.

Lamphear, B. J.; Barker, D. K.; Brooks, C. A.; Delaney, D. E.; Lane, J. R.; Beifuss, K.; Love, R.; Thompson, K.; Mayor, J.; Clough, R.; Harkey, R.; Poage, M.; Drees, C.; Horn, M. E.; Streatfield, S. J.; Nikolov, Z.; Woodard, S. L.; Hood, E. E.; Jilka, J. M.; Howard, J. A. Expression of the sweet protein brazzein in maize for production of a new commercial sweetener. *Plant Biotechnol. J.* **2005,** 3 (1), 103-114.

Liebler, D. C. Introduction to proteomics: tools for the new biology. Totowa, NJ : Humana Press. **2002**, 167-184.

Manca, C.; Lyashchenko, K.; Colangeli, R.; Gennaro, M. L. MTC28, a novel 28-kilodalton proline-rich secreted antigen specific for the Mycobacterium tuberculosis complex. *Infect. Immun.* **1997,** 65 (12), 4951-4957.

Merle, C.; Perret, S.; Lacour, T.; Jonval, V.; Hudaverdian, S.; Garrone, R.; Ruggiero, F.; Theisen, M. Hydroxylated human homotrimeric collagen I in *Agrobacterium tumefaciens*-mediated transient expression and in transgenic tobacco plant. *FEBS Lett.* **2002,** 515 (1-3), 114-118.

Miller, E. J.; Gay, S. Collagen - an Overview. *Methods Enzymol.* **1982,** 82, 3-32.

Myllyharju, J.; Nokelainen, M.; Vuorela, A.; Kivirikko, K. I. Expression of recombinant human type I-III collagens in the yeast *Pichia pastoris*. *Biochem. Soc. Trans.* **2000,** 28, 353-357.

Nandi, S.; Yalda, D.; Lu, S.; Nikolov, Z.; Misaki, R.; Fujiyama, K.; Huang, N. Process development and economic evaluation of recombinant human lactoferrin expressed in rice grain. *Transgenic Res.* **2005,** 14 (3), 237-249.

Nokelainen, M.; Tu, H. M.; Vuorela, A.; Notbohm, H.; Kivirikko, K. I.; Myllyharju, J. High-level production of human type I collagen in the yeast *Pichia pastoris*. *Yeast* **2001,** 18 (9), 797-806.

Olsen, D.; Jiang, J.; Chang, R.; Duffy, R.; Sakaguchi, M.; Leigh, S.; Lundgard, R.; Ju, J.; Buschman, F.; Truong-Le, V.; Pham, B.; Polarek, J. W. Expression and

characterization of a low molecular weight recombinant human gelatin: development of a substitute for animal-derived gelatin with superior features. *Protein Expr. Purif.* **2005,** 40 (2), 346-357.

Olsen, D.; Yang, C. L.; Bodo, M.; Chang, R.; Leigh, S.; Baez, J.; Carmichael, D.; Perala, M.; Hamalainen, E. R.; Jarvinen, M.; Polarek, J. Recombinant collagen and gelatin for drug delivery. *Adv. Drug Delivery Rev.* **2003,** 55 (12), 1547-1567.

Olsen, D. R.; Leigh, S. D.; Chang, R.; McMullin, H.; Ong, W.; Tai, E.; Chisholm, G.; Birk, D. E.; Berg, R. A.; Hitzeman, R. A.; Toman, P. D. Production of human type I collagen in yeast reveals unexpected new insights into the molecular assembly of collagen trimers. *J. Biol. Chem.* **2001,** 276 (26), 24038-24043.

Pakkanen, O.; Hamalainen, E. R.; Kivirikko, K. I.; Myllyharju, J. Assembly of stable human type I and III collagen molecules from hydroxylated recombinant chains in the yeast *Pichia pastoris* - Effect of an engineered C-terminal oligomerization domain foldon. *J. Biol. Chem.* **2003,** 278 (34), 32478-32483.

Pakkanen, O.; Pirskanen, A.; Myllyharju, J. Selective expression of nonsecreted triple-helical and secreted single-chain recombinant collagen fragments in the yeast *Pichia pastoris*. *J. Biotechnol.* **2006,** 123 (2), 248-256.

Perret, S.; Merle, C.; Bernocco, S.; Berland, P.; Garrone, R.; Hulmes, D. J. S.; Theisen, M.; Ruggiero, F. Unhydroxylated triple helical collagen I produced in transgenic plants provides new clues on the role of hydroxyproline in collagen folding and fibril formation. *J. Biol. Chem.* **2001,** 276 (47), 43693-43698.

Ruggiero, F.; Exposito, J. Y.; Bournat, P.; Gruber, V.; Perret, S.; Comte, J.; Olagnier, B.; Garrone, R.; Theisen, M. Triple helix assembly and processing of human collagen produced in transgenic tobacco plants. *FEBS Lett.* **2000,** 469 (1), 132-136.

Uchimiya, H.; Iwata, M.; Nojiri, C.; Samarajeewa, P. K.; Takamatsu, S.; Ooba, S.; Anzai, H.; Christensen, A. H.; Quail, P. H.; Toki, S. Bialaphos treatment of transgenic rice plants expressing a bar gene prevents infection by the sheath blight pathogen (*Rhizoctonia-Solani*). *Bio-Technol.* **1993,** 11 (7), 835-836.

Vuorela, A.; Myllyharju, J.; Nissi, R.; Pihlajaniemi, T.; Kivirikko, K. I. Assembly of human prolyl 4-hydroxylase and type III collagen in the yeast *Pichia pastoris*: formation of a stable enzyme tetramer requires coexpression with collagen and assembly of a stable collagen requires coexpression with prolyl 4-hydroxylase. *Embo J.* **1997,** 16 (22), 6702-6712.

Werten, M. W. T.; Wisselink, W. H.; van den Bosch, T. J. J.; de Bruin, E. C.; de Wolf, F. A. Secreted production of a custom-designed, highly hydrophilic gelatin in *Pichia pastoris*. *Protein Eng.* **2001,** 14 (6), 447-454.

Zhang, C.; Baez, J.; Glatz, C. E. Purification and characterization of a 44 kDa recombinant collagen I alpha 1 fragment from corn seed. To be submitted to *Journal of Agricultural and Food Chemistry* **2008**.

Zhong, Q. X.; Xu, L.; Zhang, C.; Glatz, C. E. Purification of recombinant aprotinin from transgenic corn germ fraction using ion exchange and hydrophobic interaction chromatography. *Appl. Microbiol. Biotechnol.* **2007,** 76 (3), 607-613.

CHAPTER 4. FRACTIONATION OF TRANSGENIC CORN SEED BY DRY AND WET MILLING TO RECOVER RECOMBINANT COLLAGENS

A paper to be submitted to Biotechnology Progress

Cheng Zhang[1], Lawrence A. Johnson[2], and Charles E. Glatz[3]

Abstract

Corn continues to be considered an ideal transgenic host for producing recombinant therapeutic and industrial proteins because of its potential for producing recombinant proteins at large volume and low cost. Efforts to reduce production costs have been primarily devoted to increasing expression level, optimizing protein extraction conditions, and simplifying purification. In the present work, we evaluated two fractionation methods, dry-milling and wet-milling, to enrich two recombinant proteins, a recombinant 100-kDa human collagen type I alpha 1(CIα1) chain and a 44 -kDa CIα1 chain fragment, for the purpose of reducing the mass and impurity burden in subsequent protein extraction and purification. For both proteins, ca. 60% of the total CIα1 was recovered in the dry-milled germ-rich fractions making up ca. 25% of the total kernel mass. For wet milling, ca. 60% of the total CIα1 was recovered in three fractions accounting for 20-25% of the total kernel mass. The CIα1s in the dry-milled germ-rich fractions were enriched 3-6 times compared to the whole corn kernel,

[1] Primary researcher and author, responsible for protein assays and data analysis
[2] Co-author, supplying technical support of corn milling processes
[3] Author for correspondence

while the CIα1s were enriched 4-10 times in selected wet-milled fractions. The recovered starch from wet milling was almost free of CIα1. Therefore, it was possible to generate CIα1-enriched fractions with both dry and wet millings and a CIα1-free starch using wet milling. Using our procedure, dry milling could be accomplished on-farm, thus minimizing the risk of inadvertent release of viable transgenic seeds.

Key words: recombinant proteins, human type I collagen, gelatin, protein extraction, transgenic corn, maize, dry milling, wet milling.

Introduction

Transgenic corn, along with a few other plants, has been extensively studied for the production of recombinant therapeutic and industrial proteins. Corn offers the advantages of low-cost cultivation; easy scale-up (increased planted acreage); well-established practices for efficient harvesting, transporting, storing and processing; and elimination of animal-source pathogens (Menkhaus, et al. 2004; Stoger, et al. 2005; Sparrow, et al. 2007). The feasibility of commercially producing high-volume (thousands of metric tons per year) and low-value (less than US$100/kg) recombinant proteins, such as collagens or gelatins, however, requires not only high accumulation level, but also efficient utilization of transgenic biomass and realization of value-added co-product revenues.

A typical approach to achieve high expression level is to target expression to a specific tissue. Advocates of germ targeting have used the embryo-preferred globulin-1 promoter; and, as a result, they have reported higher expression levels of recombinant β-

glucuronidase (GUS) (Witcher, et al. 1998), avidin (Hood, et al. 1997), fungal laccase (Hood, et al. 2003), bovine trypsin (Woodard, et al. 2003), and brazzein (Lamphear, et al. 2005) than when using constitutive or endosperm-preferred promoters. An added advantage for targeting expression to germ is the smaller amount of biomass entering downstream processing (Howard and Hood 2005) because germ is only 11% of the kernel mass (Johnson 2000). Effective milling processes for grain fractionation are required to realize this advantage and yield losses have been apparent in previous work. Recoveries in the germ-rich fractions were 55% for recombinant avidin (Hood, et al. 1997) and 60% for recombinant β-glucuronidase (Yildirim, et al. 2002). A disadvantage associated with germ targeting is that a clean germ fraction is 33% oil, which could interfere with protein extraction and separation media (Kusnadi, et al. 1998), necessitating a defatting step. Another disadvantage is that the high content of soluble germ proteins increases the separation challenge, whereas endosperm consists mostly of starch and low solubility zeins. Protein extraction from endosperm with a neutral salt buffer would leave nearly 80% of the native proteins in the spent meal (Menkhaus, et al. 2004). These are reasons why others have targeted recombinant proteins to endosperm, e.g. in the production of lipases (Zhong, et al. 2006).

Well-established dry- and wet-milling processes have been utilized by the milling industry to fractionate corn components for value-added products, but they have not been extensively studied to fractionate corn as a means to enrich recombinant proteins. Only a few studies on the dry-milling process for that purpose have been published (Hood, et al. 1997; Yildirim, et al. 2002) and no extensive study on the fractionation efficiency of the wet-milling process for the recovery of recombinant proteins could be located. In the dry-milling studies, the corn kernel was separated into anatomical fractions (germ, endosperm, and bran).

Because the majority of the oil is accumulated in the corn germ, the oil content and recovery in each fraction were most commonly used as a natural measure of the purity of and germ endosperm fractions. Targeted expression of green fluorescent protein (GFP) was recently effectively used as marker (Shepherd, et al. 2008) for fractionation quality, but it would need to be co-expressed with the desired recombinant protein to improve on the oil measure for a specific grain. Typical oil content in an industrial dry-milled germ-rich fraction for producing food-grade corn grits and flour is about 23% db (Johnston, et al. 2005) , which corresponds to about 70% purity based on the assumption that pure germ contains 33% oil (Johnson 2000). The simplified kg-scale dry milling process developed at Iowa State University does not reach this level – e.g. a recombinant dog gastric lipase (rDGL)-containing germ fraction with ca. 14% oil and 72% germ yield (Octaviani, personal communication), a recombinant β-GUS-containing germ fraction with 12% oil and 44% germ yield (Yildirim, et al. 2002), and a GFP-containing germ fraction with 17% oil and ca. 60% germ yield (Shepherd, et al. 2008). The oil contents of the endosperm-rich fractions are typically less than 1% from an effective industrial dry-milling process. In the wet-milling process, the corn kernel is separated into a greater number of components (starch, gluten/protein, fiber, germ/oil).

Compared to the dry-milled fractions, wet-milled fractions are purer and more suitable for use as ingredients or further refining (Johnson and May 2003). But the wet-milling process breaks down the kernel structure by using aqueous solutions of sulfur dioxide and lactic acid (~ pH 3.2, (Vignaux, et al. 2006)), which are detrimental to some recombinant proteins (Yildirim, et al. 2002) and could well have the negative consequence of distributing soluble proteins into multiple fractions.

Our previous work has shown that 100-kDa CIα1 and a 44-kDa fragment of CIα1 (termed 44-kDa CIα1, hereafter) can be expressed with germ targeting (Zhang, et al. 2008a; 2008b). The purpose of this work is to evaluate how well milling processes can generate germ-rich fractions enriched in biologically active recombinant protein, reducing the mass to downstream processing with the rest available to be added to the usual biorefinery fermentation feed stream.

Materials and Methods

Grain

Two transgenic corns, separately containing 100-kDa CIα1 and 44-kDa CIα1, were provided by former ProdiGene Inc (College Station, TX). Upon receipt at Iowa State University (ISU) the seeds were stored at 4^{o}C. The expression vector contained a signal sequence for export to the cell wall via the endoplasmic reticulum (ER) (Zhang, et al. 2008a); on the basis of expression levels realized this appeared to have occurred for 44-kDa CIα1, but the 100-kDa CIα1 appeared to have ended up in the ER.

Dry- and Wet-Milling Processes

Dry milling was performed using a procedure developed at the Center for Crops Utilization Research (CCUR), ISU for fractionating rDGL-containing corn (Vignaux, et al. 2004). Corn kernels were moisture-conditioned (at 21% moisture for 2.5 h), degermed, passed through roller mills and an aspirator, and fractionated through a series of sieves. A laboratory Beal-type drum degermer, which was most effective in rDGL corn fractionation,

was selected. The wet-milling process (Vignaux, et al. 2006) separates a 100-g batch into starch, gluten meal, germ, pericarp, coarse fiber and fine fiber fractions.

CIα1 Distribution in the Corn Kernel

To estimate the amount of the recombinant protein accumulated in pure germ, 100 g of each transgenic kernel was soaked in 300 ml of deionized water for 24 h. The germ, endosperm, and bran were separated by hand-dissection.

Protein Extraction and Sample Analysis

The extraction protocol was designed to recover the recombinant product with minimal co-extraction of the endogenous corn host proteins, thus reducing the complexity of the extract for subsequent purification. The milled fractions were further ground into flour using a household coffee mill (GE, Fairfield, CT). The fractions with oil contents >4% (i.e. the germ and the coarse fiber fractions and the whole corn kernel flour) were defatted before extraction. Defatting was carried out by mixing the ground corn flour with hexane (1:5 w/v) using a magnetic stir bar for 60 min at 0^{o}C. The supernatant was decanted after centrifugation (15 min, 2000g, 23^{o}C) and the extraction with hexane was repeated as before. Residual hexane in the flour was removed by air drying at room temperature. The defatted corn flour was extracted twice by mixing with extraction buffer (0.1 M phosphoric acid, 0.15 M sodium chloride, pH 1.8) in a ratio of 1:5 w/v for 1 h at room temperature. The extracts were clarified by centrifugation (~3000g, 10 min at 23^{o}C), then pooled and filtered through a 0.45μm membrane (GV Durapore PVDF, Millipore) before CIα1 and total soluble protein analysis.

The CIα1 concentrations in the extracts were quantified by a competitive ELISA (Enzyme-Linked ImmunoSorbent Assay) as previously described (Zhang, et al. 2008a; 2008b). The total protein concentration in the extract was determined by using the Pierce's Coomassie Plus - The Better Bradford™ Assay Kit (Pierce, Rockford, IL), with bovine serum albumin as the standard. All analytical assays were done in triplicate and reported as the mean.

Oil and Moisture Analyses

A Goldfisch apparatus (Labconco Corporation, Kansas City, MO) was used to determine the oil contents of ground samples by using hexanes in AACC Methods 30-26 (AACC 2000). The amounts of extracted oil were measured gravimetrically. The moisture contents of the various fractions were determined by AACC Methods 44-15A (AACC 2000).

Results and Discussion

Dry-Milling Fractionation and CIα1 Distribution

As shown in Table 1, the oil contents of the dry-milled germ-rich fractions were 10.8 and 10.3% for the 44-kDa and 100-kDa CIα1s, respectively, with estimated germ purities in those fractions being 31-33% based on the assumption of ca. 33% oil in a typical pure germ (Johnson 2000). The germ-rich fraction of the 100-kDa CIα1 captured 48% of total oil found in the corn grain, while more oil (~59% of the total oil) was recovered in the germ-rich fraction containing the 44-kDa CIα1. In contrast, the endosperm-rich and bran-rich fractions from both 44-kDa and 100-kDa CIα1 corn grain have relatively high oil content (~3%)

Table 1. Oil contents in fractions produced by dry milling

Corn Type	Fraction Mass Yield (db %)		Oil (db %)	Total Oil (%)
44-kDa CIα1	Germ	27	10.8	59
	Endosperm	67	2.7	38
	Bran	5	3.1	3
100-kDa CIα1	Germ	22	10.3	48
	Endosperm	69	3.3	49
	Bran	6	2.6	3

compared to that reported in literature (~1% oil in both fractions) (Johnson 2000), as a result of contamination with the remaining germ.

The 31-33% germ purity obtained in our 100-kDa CIα1 and 44-kDa CIα1 containing germ-rich fractions was close to what has been achieved in dry-milling other transgenic corn, such as rDGL (Octaviani, personal communication) and a β-GUS-containing germ fraction (Yildirim, et al. 2002). Theoretically, proteins targeted to germ can be enriched 8- to 9-fold if a pure germ fraction can be obtained because germ only accounts for 11-12% of the total kernel mass, but the 31-33% germ purity achieved by the dry milling in this work indicated that germ was only enriched 3-fold with 50-60% recovery.

The purity of the germ was increased by a modified milling procedure that passed the 44-kDa CIα1-containing germ-rich fraction through one additional milling and sieving operation to generate a high purity (>60% germ) germ-rich fraction. This procedure enriched the already concentrated CIα1 three times to 120 mg/kg of dry solids, but the improvement was at the expense of lower germ (ca. 40%) and CIα1 (ca. 35%) yields.

The CIα1 concentration and yield in the hand-dissected germ fraction provided an upper limit for judging the dry-milling process that eliminates germ loss and accounts for any

"leaky" expression from the embryo-preferred globin-1 promoter to endosperm. The hand-dissected germs contained ca. 34% oil with the corresponding endosperm containing 0.7% oil, levels are consistent with the literature (Johnson 2000). As shown in Table 2, although only 67% of total 100-kDa CIα1was located in the pure germ, it was concentrated 18 times (42.2 μg CIα1/g of dry solids) compared to that in endosperm (2.3 μg CIα1/g of dry solids). Similarly, 62% of total 44-kDa CIα1was accumulated in the germ with concentration (175.8 μg CIα1/g of dry solids) as much as 13 times that in endosperm (13.4 μg CIα1/g of dry solids)**.**

The dry-milling processes recovered 63% of the total 100-kDa CIα1 in a germ-rich fraction comprised of 22% of the total kernel mass (Table 3). This recovery was higher than that of oil recovery (Table 1), which might be anticipated because endosperm contamination of this fraction carries along more CIα1 than oil. However, for the 44-kDa CIα1 the recovery closely mirrored that of the oil (Tables 1 and 4). The difference could result from the 15-20% variation in the CIα1 ELISA assay. The 100-kDa CIα1 was 5.5 times enriched in the germ-rich fraction (15.9 μg/g dry solids) compared to that in the whole corn kernel (2.9 μg/g dry solids). The 44-kDa CIα1 in the germ-rich fraction was 2.6-fold concentrated (49.6 μg/g dry solids) relative to the whole kernel (19.2 μg/g dry solids). Most of the rest of the 44-kDa CIα1 ended up in the endosperm-rich fraction. In general, dry-milling processes were capable of generating a CIα1-enriched fraction with CIα1 concentration close to one-half that in the pure germ fractions at ca. 60% CIα1 yield. It should be noted that the CIα1 yield of each fraction was calculated on the basis of the summation of CIα1s from all fractions, rather than the lower total CIα1s obtained from whole corn extraction (Table 3-6 footnotes). All dry fractions received a final milling by using a coffee grinder and the same extraction conditions

Table 2. CIα1 distributions by hand-dissection

Corn Type	Fraction Mass Yield (db %)		CIα1		
			µg/g dry solids	Mass Yield (%)	Purity in Extract (%)*
44-kDa CIα1	Germ	11	175.8	62	1.94
	Endosperm	80	13.4	34	0.29
	Bran	9	14.9	4	0.22
100-kDa CIα1	Germ	11	42.2	67	0.41
	Endosperm	86	2.3	30	0.05
	Bran	4	5.3	3	0.07

***** Purity in extract is defined as the concentration of the CIα1 divided by the total protein extracted under the extraction condition (0.1 M phosphoric acid, 0.15 M sodium chloride, pH 1.8)

in an effort to equalize the expected extractabilities. Nonetheless, the milled fractions had been previously submitted to multiple steps that could have increased the extraction yield and for the wet-milled liquid fractions the extractant was different.

These germ-fraction recoveries were consistent with previously-referenced dry-milling studies showing ca. 60% recovery. In contrast, close to 90% of rDGL was recovered in the dry-milled endosperm (Octaviani, personal communication). Because endosperm accounts for >80% of the kernel mass, the fraction of endosperm lost to the smaller germ and bran fractions was less, an inherent advantage of endosperm expression that offsets the lower biomass burden reduction. In addition, oil content was lower in the endosperm fraction and the amount of native corn proteins extracted from corn endosperm remained constantly low in the pH range 4-9 while the host protein content of pH 9 germ extracts was ca. 25 times more than that of pH 4 extracts (Gu and Glatz 2007). If expression levels comparable to those in germ could be reached in endosperm, then endosperm could be more appropriate for protein expression especially when the recombinant protein needs to be extracted at neutral

Table 3. 100-kDa CIα1 distributions by the dry-milling process

Fraction Mass Yield (db %)		CIα1			TEP*
		µg/g dry solids	Mass Yield (%)**	Purity in Extract (%)	Mass Yield (%)
Germ	22	15.9	63	0.29	23
Endosperm	69	2.6	33	0.05	71
Bran	6	3.5	4	0.07	6
Whole kernel	--	2.9	--	0.05	--

* TEP: total extractable proteins in the acidic extraction condition

** Mass yield is based on the total summation of CIα1 from all fractions (480 µg CIα1/100 g corn) rather than whole corn without fractionation (260 µg/100 g corn)

Table 4. 44-kDa CIα1 distributions by the dry-milling process

Fraction Mass Yield (db %)		CIα1			TEP
		µg/g dry solids	Mass Yield (%)*	Purity in Extract (%)	Mass Yield (%)
Germ	27	49.6	55	0.93	30
Endosperm	67	14.5	40	0.31	65
Bran	5	23.9	5	0.47	5
Whole kernel	--	19.2	--	0.37	--

* Mass yield is based on the total summation of CIα1 from all fractions (2070 µg/100 g corn) rather than whole corn without fractionation (1650 µg/100 g corn)

or alkaline pH.

CIα1 Distribution in Wet-Milled Fractions

Some similarities were observed between the two collagens. As shown in Tables 5 and 6, both the 44-kDa CIα1 and the 100-kDa CIα1were present in all wet-milled fractions except for starch. About 65% of the total 100-kDa CIα1 found in the corn kernel was recovered in the germ, coarse fiber, and fine fiber fractions comprising 17% of the final dry mass (Table 7). The 100-kDa CIα1 concentrations in those fractions were 4-6 times higher than in the whole kernel. Similarly, 63% of the total 44-kDa CIα1was captured in the fine fiber, coarse fiber, and gluten meal (as opposed to germ) fractions, which accounted for 26% of the final dry mass. The 44-kDa CIα1 contents in both fiber fractions were concentrated about 10-fold compared to the whole kernel. Although both collagens were targeted to the germ, they were not confined in the germ fractions after wet milling. Only 3% of the 44-kDa and 22% of the 100-kDa CIα1 were recovered in the germ fractions. Loss from the germ is not surprising because the CIα1s are soluble at the steeping conditions (and significant CIα1 is in the liquid fractions); however, it is not clear what would cause their association with other solid fractions. Targeting to the cell wall would give an initial cell wall association and the fine fiber contains cell walls. The coarse fiber (primarily bran) could still be attached to a portion of the germ. The suspected retention of the 100-kDa CIα1 in the ER could explain its greater retention in the germ (22 vs. 3%), which also suggests differences in localization offer the potential to influence distribution, particularly during wet milling. Still higher subcellular entrapment would be necessary for targeting effectiveness to be maintained through wet milling. The recovery of protein-free starch facilitates its use for fermentation

Table 5. 100-kDa CIα1 distribution in the wet-milling process

Fraction Mass Yield (db %)		CIα1			TEP
		µg/g dry solids	Mass Yield (%)*	Purity in Extract (%)	Mass Yield (%)
Starch	58	0.1	1	-	0
Gluten Meal	12	2.1	7	0.13	7
Germ	6	12.7	22	0.18	16
Pericarp	4	4.5	5	0.14	5
Coarse Fiber	5	18.8	25	0.24	14
Fine Fiber	6	11.5	18	0.27	9
Steep Water	6	-	12	0.07	21
Rinse Water	3	-	9	0.04	29

***** Mass yield was based on the total summation of CIα1 from all fractions (302 µg/ 100 g corn) rather than whole corn without fractionation (260 µg/100 g corn)

Table 6. 44-kDa CIα1 distribution in the wet-milling process

Fraction Percent (db %)		CIα1			TEP
		µg/g dry solids	Mass yield (%)*	Purity in extract (%)	Mass yield (%)
Starch	56	0.1	0	0.9	0
Gluten Meal	14	20.5	8	2.4	2
Germ	2	44.0	3	0.5	4
Pericarp	3	31.3	3	1.2	2
Coarse Fiber	6	211.0	34	2.6	9
Fine Fiber	6	111.3	21	4.0	3
Steep Water	6	-	12	0.5	17
Rinse Water	5	-	19	0.2	64

***** Mass yield was based on the total summation of CIα1 from all fractions (2912 µg/100 g corn) rather than whole corn without fractionation (1650 µg/100 g corn)

Table 7. Comparison of dry and wet milling (data deriving from Table 3-6)

Process	Product Fractions	Remaining Solids (%)	CIα1 Yield (%)	CIα1 Conc. (μg/g dry solids)	Purity in Extract (%)	PF*
44-kD Wet	Fine/Coarse fiber/Gluten	26	63	84.6	2.84	7.7
44-kD Wet	Fine/Coarse fiber only	12	55	157.8	3.32	9.0
44-kD dry	Germ	27	55	49.6	0.93	2.5
100-kDa Wet	Fine/Coarse fiber/Germ	17	65	14.0	0.23	4.6
100-kDa dry	Germ	22	63	15.9	0.29	5.8

* PF: purification factor, defined as the CIα1 purity in the extract from the fraction/ fractions divided by the corresponding CIα1 purity from the whole corn (without fractionation)

(such as fuel ethanol).

In the two published reports of recombinant protein partitioning during wet milling, Vignaux et al. (Vignaux and Johnson 2006) reported that the *Escherichia coli* heat-labile enterotoxin B subunit (LT-B) expressed in corn endosperm was also concentrated in fiber fractions. About 85% of glucuronidase activity was recovered in a germ fraction containing 10% of dry mass when steeping without SO_2 (Johnson and May 2003), but the SO_2-free steep alternative did not improve the outcome (results not shown).

Despite dispersing the CIα1s, both milling processes can increase the CIα1 purity in selected solid fractions. Concentration in the fiber fractions makes wet milling a feasible alternative; it is actually better (by measures of purity and purification factor) than the dry-milling alternative (Table 7) though the latter's performance could be improved by further processing to improve germ purity.

While the main advantage offered by fractionation is the reduction in volume and

impurity load to the separation train, an initial on-farm dry-milling step would eliminate the need to transport viable transgenic seeds, thus reducing risk of accidental release.

Conclusions

Both dry milling and wet milling can be utilized to enrich the CIα1s and reduce the burden to subsequent protein extraction and purification. Dry milling allowed 50-60% of the total CIα1s to be captured in a germ-rich fraction making up ca. 25% of the total kernel mass. The recombinant protein was enriched 2.5 and 5.8 times for 44-kDa CIα1 and 100-kDa CIα1, respectively.

For the wet milling, 63% of total 44-kDa CIα1 was captured in fine fiber-, coarse fiber-, and gluten-rich fractions which accounted for 26% of the total kernel mass, and 65% of total 100-kDa CIα1 was recovered in fine fiber-, coarse fiber-, and germ-rich fractions, which made up 17% of the total kernel mass. The recombinant protein was enriched 7.7 and 4.6 times for 44-kDa CIα1 and 100-kDa CIα1, respectively. If only considering the fiber fractions, 55% of total 44-kDa CIα1 was captured in 12% of total kernel mass, with 9-fold enrichment, which represents a big potential for reducing the steps of the subsequent purification.

While for the 100-kDa CIα1, wet and dry milling showed similar performance, for the 44-kDa CIα1, wet milling is superior to dry milling because of greater CIα1 purity in the extract from the combined fiber- and gluten-rich fractions and higher purification factor (PF). The mechanism behind the association of the CIα1 with both coarse fiber and fine fiber observed in this study, as well as in the case of LT-B, was not clear. In wet milling, transport

via solution distributes the product to more fractions, but the clean separation of the starch free of proteins and oil was an advantage. Dry milling may be easier to incorporate into the recombinant protein production because of its simple operation, low cost and partial protein enrichment. In conclusion, either milling processes could be a feasible prefractionation step for production of recombinant proteins.

Acknowledgements

This work was supported by the funding from the USDA CREES Grant #2006-34496-17122. The authors would like to thank Julio Baez from FibroGen for providing technical review of this work, Steve Fox for providing support in the milling processes and Janice Seibel for the analysis of the 44-kDa CIα1 wet-milled fractions.

References

AACC, Approved Methods of the American Association of Cereal Chemists, 10th Ed. St. Paul, MN. 2000.

Gu, Z. R.; Glatz, C. E. Aqueous two-phase extraction for protein recovery from corn extracts. *J. Chromatogr. B* **2007,** 845 (1), 38-50.

Hood, E. E.; Bailey, M. R.; Beifuss, K.; Magallanes-Lundback, M.; Horn, M. E.; Callaway, E.; Drees, C.; Delaney, D. E.; Clough, R.; Howard, J. A. Criteria for high-level expression of a fungal laccase gene in transgenic maize. *Plant Biotechnol. J.* **2003,** 1 (2), 129-140.

Hood, E. E.; Witcher, D. R.; Maddock, S.; Meyer, T.; Baszczynski, C.; Bailey, M.; Flynn, P.; Register, J.; Marshall, L.; Bond, D.; Kulisek, E.; Kusnadi, A.; Evangelista, R.; Nikolov, Z.; Wooge, C.; Mehigh, R. J.; Hernan, R.; Kappel, W. K.; Ritland, D.; Li, C. P.; Howard, J. A. Commercial production of avidin from transgenic maize: characterization of transformant, production, processing, extraction and purification. *Mol. Breed.* **1997,** 3 (4), 291-306.

Howard, J. A.; Hood, E. Bioindustrial and biopharmaceutical products produced in plants. *Adv. Agron.* **2005,** 85, 91-124.

Johnson, L. A., Corn: The major cereal of the Americas. In *Handbook of Cereal Science and Technology*, 2nd ed.; Kulp, K.; Ponte, J. G., Eds. New York, 2000; pp 31-80.

Johnson, L. A.; May, J. B., Wet milling: The basis for corn biorefineries. In *Corn: Chemistry and Technology*, 2nd ed.; White, P. J.; Johnson, L. A., Eds. American Association of Cereal Chemists, Inc: St. Paul, MN, 2003; pp 449-494.

Johnston, D. B.; McAloon, A. J.; Moreau, R. A.; Hicks, K. B.; Singh, V. Composition and economic comparison of germ fractions from modified corn processing technologies. *J. Am. Oil Chem. Soc.* **2005,** 82 (8), 603-608.

Kusnadi, A. R.; Evangelista, R. L.; Hood, E. E.; Howard, J. A.; Nikolov, Z. L. Processing of transgenic corn seed and its effect on the recovery of recombinant beta-glucuronidase. *Biotechnol. Bioeng.* **1998,** 60 (1), 44-52.

Lamphear, B. J.; Barker, D. K.; Brooks, C. A.; Delaney, D. E.; Lane, J. R.; Beifuss, K.; Love, R.; Thompson, K.; Mayor, J.; Clough, R.; Harkey, R.; Poage, M.; Drees, C.; Horn, M. E.; Streatfield, S. J.; Nikolov, Z.; Woodard, S. L.; Hood, E. E.; Jilka, J. M.; Howard, J. A. Expression of the sweet protein brazzein in maize for production of a new commercial sweetener. *Plant Biotechnol. J.* **2005,** 3 (1), 103-114.

Menkhaus, T. J.; Bai, Y.; Zhang, C. M.; Nikolov, Z. L.; Glatz, C. E. Considerations for the recovery of recombinant proteins from plants. *Biotechnol. Prog.* **2004,** 20 (4), 1001-1014.

Shepherd, C. T.; Vignaux, N.; Peterson, J. M.; Scott, M. P.; Johnson, L. A. Dry-milling and fractionation of transgenic maize seed tissues with green fluorescent protein as a tissue marker. *Cereal Chem.* **2008,** 85 (2), 196-201.

Sparrow, P. A. C.; Irwin, J. A.; Dale, P. J.; Twyman, R. M.; Ma, J. K. C. Pharma-Planta: Road testing the developing regulatory guidelines for plant-made pharmaceuticals. *Transgenic Res.* **2007,** 16 (2), 147-161.

Stoger, E.; Ma, J. K. C.; Fischer, R.; Christou, P. Sowing the seeds of success: pharmaceutical proteins from plants. *Curr. Opin. Biotechnol.* **2005,** 16 (2), 167-173.

Vignaux, N.; Fox, S. R.; Johnson, L. A. A 10-g laboratory wet-milling procedure for maize and comparison with larger scale laboratory procedures. *Cereal Chem.* **2006,** 83 (5), 482-490.

Vignaux, N.; Johnson, L. A., Wet Milling: an efficient process to recover a recombinant LTB-rich fraction from corn. In *Annual Meeting of American Association of Cereal Chemists*, Program Book, World Grains Summit: Foods and Beverages, Abstract P322, pp 173: San Francisco, CA. Sept. 17-20, 2006.

Vignaux, N.; Octaviani, D.; Johnson, L. A., Efficiencies of different types of dry mills in recovering a fraction rich in recombinant protein expressed in endosperm. In *Annual Meeting of American Association of Cereal Chemists and the Tortilla Industry Association*, AACC/TIA Annual Meeting, Program Book,Vol Abstract 305, p141: San Diego, CA. Sept 19-22, 2004.

Witcher, D. R.; Hood, E. E.; Peterson, D.; Bailey, M.; Bond, D.; Kusnadi, A.; Evangelista, R.; Nikolov, Z.; Wooge, C.; Mehigh, R.; Kappe, W.; Register, J.; Howard, J. A. Commercial production of beta-glucuronidase (GUS): a model system for the production of proteins in plants. *Mol. Breed.* **1998,** 4 (4), 301-312.

Woodard, S. L.; Mayor, J. M.; Bailey, M. R.; Barker, D. K.; Love, R. T.; Lane, J. R.; Delaney, D. E.; McComas-Wagner, J. M.; Mallubhotla, H. D.; Hood, E. E.; Dangott, L. J.; Tichy, S. E.; Howard, J. A. Maize (Zea mays)-derived bovine trypsin: characterization of the first large-scale, commercial protein product from transgenic plants. *Biotechnol. Appl. Biochem.* **2003,** 38, 123-130.

Yildirim, S.; Fuentes, R. G.; Evangelista, R.; Nikolov, Z. L. Fractionation of transgenic corn for recovery of recombinant enzymes. *J. Am. Oil Chem. Soc.* **2002,** 79 (8), 809-814.

Zhang, C.; Baez, J.; Glatz, C. E. Purification and characterization of a 44 kDa recombinant collagen I alpha 1 fragment from corn seed. *To be submitted to Journal of Agricultural and Food Chemistry* **2008a**.

Zhang, C.; Baez, J.; Glatz, C. E. Purification and characterization of recombinant collagen type I alpha 1 from transgenic corn seed. *To be submitted to Biotechnology Progress* **2008b**.

Zhong, Q. X.; Gu, Z. R.; Glatz, C. E. Extraction of recombinant dog gastric lipase from transgenic corn seed. *J. Agric. Food Chem.* **2006,** 54 (21), 8086-8092.

CHAPTER 5. GENERAL CONCLUSIONS AND FUTURE WORK

General Conclusions

A purification process has been developed without affinity chromatography to yield a pure 100 kDa CIα1 with 16% recovery from an extremely low expression transgenic source. The purification process included one step of membrane filtration plus four steps of chromatographic columns. Purification of the 44 kDa CIα1 benefited from the source transgenic grain having 6 times higher expression level than for the 100 kDa CIα1 that also made it worthwhile to separate the grain into a germ rich fraction with still higher levels. The result was purification in only two chromatographic steps.

Characterization of the purified 44 kDa CIα1 demonstrated that 1) its intact protein mass measured by MS was within 0.2% of the mass calculated from the expected amino acid sequence, 2) 78% of its primary sequence was confirmed by tandem MS, 3) the first 20 amino acid residues at the N-terminus determined by N-terminal sequencing matched that expected from the gene sequence, indicating the signal sequence was properly cleaved, 4) it was immunoreactive to an anti-25 kDa CIα1 antibody, and 5) it did not show detectable glycosylation. All those data provided reliable evidence that the corn-derived 44 kDa CIα1 was consistent with the protein expressed from the recombinant DNA construct transformed into the corn plants and with only the intended post-translational modification. In other words, the results strongly indicated that corn plants were able to correctly express and process this single-stranded collagen fragment.

The transgenic corn showed less faithful expression of the longer, triple-helix version – the 100 kDa CIα1 collagen. Only a fraction of the corn-derived CIα1 was recovered with triple-helical structure. The relatively low melting temperature of the 100 kDa CIα1 (26^{o}C vs 40^{o}C in native human collagen) was well correlated with its low hydroxyproline content (low hydroxyproline also evidenced in 44 kDa CIα1), indicating both some ability of corn to hydroxylate proline, but also the necessity of co-expression of a heterologous P4H if higher content is desired. Not unexpectedly, the C-terminal foldon was not removed, so an additional processing step of pepsin digestion would be needed to convert this procollagen form to mature collagen.

However, the corn host did demonstrate several capabilities for producing collagen that go beyond the requirement of previous globular recombinant proteins expressed in corn. The purified 100 kDa CIα1 molecule was immunoreactive to an anti-collagen antibody, had a very similar molecular weight to that of an analogous CIα1 expressed in *Pichia* yeast, and 46% of its primary sequence was resolved by tandem MS, confirming its identity. With some further optimization of gene construct and transformation (e.g. co-expression of P4H, more effective promoter), the plant expression system might be able to produce higher levels of well-processed collagens with higher melting temperature.

Both dry and wet milling were capable of enriching the CIα1s to reduce biomass for the subsequent protein extraction and improve the efficiency of the purification. The 44 kDa CIα1 and 100 kDa CIα1 were enriched 2.5 and 5.8 times in the dry-milled germ-rich fractions, respectively, while they were enriched 7.7 and 4.6 times in the selected wet-milling fractions, respectively. Although wet milling showed higher enrichment, CIα1s were dispersing in several fractions. A major portion of CIα1s were unexpectedly captured in

fiber-rich fractions, but the high CIα1 concentration in relatively small amount of fiber fractions (only accounting for 11-12% of total seed biomass) makes wet milling a promising method for CIα1 enrichment. The separation of starch free of proteins and oil by wet milling was another advantage. An initial on-farm dry-milling step would offer the additional advantage of eliminating the need to transport viable transgenic seed, thus reducing risk of accidental release.

Future Work

Although the corn plant expression system has demonstrated its potential, several hurdles have to be resolved for the corn to be a viable host for CIα1s expression.

First of all, the expression level must improve hundreds of times, from current 3 mg of 100 kDa CIα1/ kg seeds to >1g of CIα1/ kg seeds, the accumulation level required by the commercial production. This high expression level is necessitated by the production cost. As low- or intermediate-valued recombinant proteins, the inexpensive, non-chromatographic methods are required, and reasonable expression level is necessary for methods such as precipitation to work effectively and reduce the purification cost.

Second, the relatively low melting temperature observed in the current corn-derived 100 kDa CIα1certainly limits its use in many applications such as pharmaceutical capsules. Therefore, co-expression of a heterologous P4H is necessary to achieve higher hydroxyproline content to form more stable triple-helical structure and increased melting temperature. By controlling the co-expression of P4H with CIα1, the level of proline hydroxylation could be tailored to suit different applications.

Third, fragments of 100 kDa CIα1 were observed during its purification, but it is uncertain whether the fragments were generated during the growth of the corn seeds, grain storage, or the extraction and purification process. Measures must be taken to reduce the fragments. If the fragments were generated during the lengthy multistep purification process by concomitantly released plant proteases from extraction, inhibitors could be added during the protein extraction. If during the synthesis of the CIα1 in seeds, gene modification might be needed. Again, higher expression in future generations would simplify the purification process and perhaps reduce the proteolytic cleavage that may have occurred during this lengthy purification process.

Fourth, we observed the unexpected association of CIα1s with fibers during milling process, but the mechanism behind the association of CIα1s with fibers was unclear. The reason why the CIα1s were captured by fibers needs to be examined as there is potential for exploiting this as a means to enrichment. Electrostatic interaction is hypothesized to be involved, because the CIα1s (pI 8-9.5) are positively charged under the steeping condition (~pH 3.2). Zeta potential of the wet fibers during milling could be measured to determine if they possess negative charge that would support this hypothesis.

ACKNOWLEDGEMENTS

I would like to extend my deep appreciation to Dr. Charles Glatz, who constantly provided advice and support throughout my graduate study. The completion of this work would not be possible without his continual guidance and encouragement. Grateful acknowledgement is extended to Drs. Julio Baez and Louisa Tabatabai for providing advice in this research. I would also like to take this opportunity to thank all members of my POS committee, Drs. Lawrence Johnson, Patricia Murphy, Balaji Narasimhan, and Peter Reilly, for their helpful reviews and support of this manuscript.

I cannot reflect upon my stay at ISU without thinking of those who helped me around. I extend a sincere thank you to Ramon Morales-Chabrand, Kerry Campbell, Matt Aspelund, Zhengrong Gu, Jason Haase, and Yi Zhang. Additional acknowledgement is extended to Steve Fox, Janice Seibel, and Siquan Luo for their technical assistance in my research.

I would like to thank my family for their everlasting love and undaunted support. Finally and most importantly, I would like to extend heartfelt thanks to my wife Gan Jin for her love, support, and, most of all, understanding.