Human Recombinant Type I Collagen Produced in Plants

Oded Shoseyov, PhD1,2 Yehudit Posen, PhD3 and Frida Grynspan, PhD2

As a central element of the extracellular matrix, collagen is intimately involved in tissue development, remodeling, and repair and confers high tensile strength to tissues. Numerous medical applications, particularly, wound healing, cell therapy, bone reconstruction, and cosmetic technologies, rely on its supportive and healing qualities. Its synthesis and assembly require a multitude of genes and post-translational modifications, where even minor deviations can be deleterious or even fatal. Historically, collagen was always extracted from animal and human cadaver sources, but bare risk of contamination and allergenicity and was subjected to harsh purification conditions resulting in irreversible modifications impeding its biofunctionality. In parallel, the highly complex and stringent post-translational processing of collagen, prerequisite of its viability and proper functioning, sets significant limitations on recombinant expression systems. A tobacco plant expression platform has been recruited to effectively express human collagen, along with three modifying enzymes, critical to collagen maturation. The plant extracted recombinant human collagen type I forms thermally stable helical structures, fibrillates, and demonstrates bioactivity resembling that of native collagen. Deployment of the highly versatile plant-based biofactory can be leveraged toward mass, rapid, and low-cost production of a wide variety of recombinant proteins. As in the case of collagen, proper planning can bypass plant-related limitations, to yield products structurally and functionally identical to their native counterparts.

Collagen Structure and Function

Collagen, secreted by fibroblasts and epithelial cells, constitutes the most dominant protein of the extracellular matrix (ECM) and connective tissue, and is intimately involved in tissue development, remodeling, repair, and overall physical support. The protein confers high-tensile strength and resilience to tissues, where its profile of post-translational modifications dictates its physical specifications and role at each deposition site.1 As the most abundant animal protein, collagen can be found in 27 variations, yet, types I, II, and III make up 80%–90% of bodily collagens.2 The soluble, 300 nm long and 1.5 nm wide macromolecule contains over 300 Gly-X-Y triplet repeats along its chain. Upon orderly packing, three collagen polypeptide chains assemble into a characteristic collagen triple helix to form sheet structures or homo- or heterotrimeric fibrils (30–300 nm wide), displaying a staggered pattern at an axial periodicity of 67 nm. The remarkable abundance of glycine, proline, and hydroxyproline, enables a stable coiled-coil conformation of the two identical α1 chains together with one α2 chain. Lateral interactions between chains, in the form of hydrogen bonds or aldol cross-links, among others, further stabilize collagen packing and strengthen fibrils. Fibril-forming collagens are flanked by terminal propeptides, which lack the Gly-Pro-Xaa motif that prevents premature fibrilization. The individuality of each collagen type is defined by the final ensemble of domain types. These variances highly influence the orientation behavior, eventual fibril diameter, and the molecular structures formed upon polymerization and their biomechanical specifications.

Type I Collagen Biosynthesis

Orchestration and control of collagen synthesis, trafficking, fibril assembly, and deposition in the extracellular space require a multitude of genes, processing enzymes, associated proteins, and cofactors (reviewed in Canty and Kadler3 and Myllyharju and Kivirikko4). To date, 42 collagen-encoding genes have been identified in the human genome. The three Type I procollagen polypeptides are encoded by the COL1A1 and COL1A2 genes, located on chromosomes 17q21-22 and 7q21-22, respectively. Their transcription is differentially regulated by cytokines, typically of inflammatory nature, mediated by the Janus Kinase family of protein tyrosine kinases and by signal transducers and activators of transcription (STATs).5,6 Upon their activation and nuclear translocation, STATs bind to responsive promoter elements, but still require

1The Robert H. Smith Faculty of Agriculture, Food and Environment, The Robert H. Smith Institute of Plant Science and Genetics, The Hebrew University of Jerusalem, Rehovot, Israel.
2CollPlant Ltd., Ness-Ziona, Israel.
3Scintax, Rehovot, Israel.

1527
the coordinated participation of a range of transcription factors. Polypeptide collagen chains are synthesized on membrane-bound ribosomes, and they are cotranslationally translocated in the endoplasmic reticulum (ER).

Post-Translation Modifications of Procollagen

The highly complex and hierarchical post-translational processing of collagen is a prerequisite of full maturation and functionality of the protein and its fibrillar product. The series of post-translational modifications of procollagen enables triple helix nucleation, formation, and protein folding in the ER. Collagen-specific hydroxylases, glycosyltransferases, proteinases, and one oxidase are required for the initial modifications of procollagen en route to collagen maturation. Activity of the multienzyme prolyl 4-hydroxylases (P4H) complex, acting on over 100 proline residues along the collagen polypeptide chain, has been implicated in directional helical conformation of collagen and in increasing the protein’s thermal stability and viability. Upon P4H deficiency or inactivity, collagen secretion is fully prevented and intermediate, partially folded procollagen chains accumulate within the ER, initiating ER stress and unfolded protein responses that contribute to the pathophysiology of inherited ECM disorders.

While the extent of lysine hydroxylation among tissue collagens is highly variable, when compared with that of proline, it plays a central role in collagen fibrillogenesis and ECM mineralization. Three isoforms of lysyl hydroxylase (LH1–3), along with two alternatively spliced LH2 variants, undergo differential expression between tissues and developmental stages and demonstrate specificity to particular domains of the collagen molecules. The hydroxylated lysine residues act as the single target of glycosylation on collagen molecules, and as sites of oxidative deamination and covalent cross-linking, critical in fibril stabilization. Hydroxyllysine glycosylation of the nascent collagen peptide, mediated by β(1→O)-galactosyl- and α(1→2) glucosyltransferase enzymes and found in marked variability among collagens, has been suggested to impact the rate of interaction with procollagen proteinases, and the extent of collagen supramolecular aggregation.

Peptidyl proline cis-trans isomerase and protein disulfide isomerase catalyze intra- and interchain disulfide bonding, in addition to the role of the former as a β-subunit of collagen P4H, and the role of the latter in triple-helix folding. In addition, the majority of procollagen N and C propeptides are cleaved by proteases located in transport vesicles, allowing for spontaneous self-assembly of collagen fibrils. Proangiogenic, anti-angiogenic, and chemotactic functions have been assigned to these cleaved, circulating propeptides, depending on their sources and on the molecular context in which they accumulate.

The sequence of events leading up to collagen maturation and assembly must proceed systematically and in its entirety, to avoid defects in their structure and/or function. The stringent structural and post-translational modification requirements of collagens render them highly susceptible to mutations; even minor defects in collagen structure or assembly can have severe clinical consequences due to the protein’s widespread distribution throughout the human body. Such abnormalities can be manifested by tissue pathology, such as impaired vasculature, tendons, gums, hair, and skin. The functional relevance of proper collagen triple helix configuration is clinically expressed upon impaired collagen hydroxylation in scurvy and in osteogenesis imperfecta, associated with severe bone fragility. Hypochondroplasia, one of the most severe disorders caused by mutations in COL2A1, is characterized by a short body and limbs, along with abnormal bone formation in the spine and pelvis. The inherited Ehlers-Danlos type VII connective tissue disease is caused by inactive ADAMTS2, the enzyme responsible for collagen type I, II, and III propeptide excision. The defective collagen molecules accumulate and self-assemble into abnormal fibrils, which alter tissue strength and structure.

Formation of Nano- and Micro-Collagen-Based Structures in the Extracellular Space

Decorated collagen fibrils are transported to the plasma membrane (PM) within Golgi-to-PM transport compartments, which then fuse with the PM to form a fibripositor. Collagen bundles and higher-order structures are assembled in these extracytoplasmic recesses, where they can linearly and laterally expand and/or form composite filaments with other types of collagen or noncollagenous elements. The fibrils then undergo a number of modifying and reinforcement processes, including oxidation and cross-linking. Tip-to-tip fusion of collagen fibril intermediates is promoted by surface-bound small leucine-rich proteoglycans, which inhibit side-to-side fibril aggregation. Post-depositional cross-linking and strengthening then takes place within the interstitial space. The resulting products preserve an intimate association with the cell surface and lay the ground for the extensive and complex tissue-specific extracellular network.

Collagens in Biomedical and Pharmaceutical Applications

Collagens boast a long history of medical and industrial use in healing procedures, tissue reconstruction, and as dermal fillers, where their abundance, biocompatibility, biodegradability, and functionality render them ideal scaffold materials for human use. Collagen-based scaffolds provide advantageous features for both cell-free and cell-based tissue engineering techniques, defining the space of tissue ingrowth, and serving as tissue/organ equivalents, respectively. Such scaffolds can be customized to attract specific cell types or impregnated with factors to be released at the graft site. Collagen scaffolds designed to closely mimic natural ECM environs, are widely applied in basic research to shed light on cell behavior and disease etiology and pathogenesis.

Specific collagen types are closely associated with stem and progenitor cell niches and cell clustering, which play a critical role in cell self-renewal, differentiation, and function. These interactions are highly dependent on the triple helical FGFR collagen sequence, the protein’s fibrillar status, and on integrin binding. These microenvironments can be reconstituted in vitro, and form the basis of cell therapies toward accelerated reconstruction of diseased tissues. Collagen-based biomaterials in the form of heart valves and functional veins have been particularly valuable in cardiac and pulmonary repair ensuing vascular
disease-related insult. Similarly, researchers are actively pursuing development of bioengineered ligament and tendon tissue mimics that can provide mechanical support until neotissue is established.\textsuperscript{35}

Throughout the multi-stage wound healing process, collagen and collagen-derived fragments provide indispensable support for cell aggregation and adhesion, clot formation, fibroblast recruitment, and adequate scar tissue generation.\textsuperscript{36} Thus, biocompatible collagen-based wound dressings contribute local haemostatic and chemotactic stimuli, while supplying a structural support upon which neotissue can be formed at enhanced rates. In addition, the highly absorptive character of such products accommodates the high exudate volume characteristic of injured tissue, preserving appropriate moisture balance of the wound bed. Collagen finds extensive use in cosmetic procedures, in the form of dermal fillers injected to smooth the overall appearance of the skin or in more aggressive reconstructive surgery. Advancements in fabrication of collagen-based corneae have been made as well\textsuperscript{37} and may eventually obviate the need for human cornea donations.

**Expression Systems for Production of Recombinant Collagen**

Historically, collagen used for pharmaceutical and medical applications was extracted from animal or cadaver sources. However, the age and physiology of the tissue from which collagen is harvested significantly impact the biochemical profile of the extracted collagen, as it is continuously subject to postdepositional modifications throughout its lifecycle. Further, tissue-derived collagens inevitably contain growth factors and/or cytokines contaminants, which can highly influence matrix–cell interactions and overall tissue remodeling prospects. Moreover, extraction protocols inflict irreversible cross-links within the protein, jeopardizing the biological and mechanical functions of collagen. In addition, growing concerns relating to their allergenicity and risk of prion/pathogen transmission have spurred research in pursuit of alternative solutions.\textsuperscript{38,39} Further accentuation of the rising need for safe biomaterials alternatives has been provided by a recent FDA publication calling for prohibition of certain cattle materials in the manufacturing of biologics and medical devices intended for humans.\textsuperscript{40} Fish-derived collagens, extensively used in food, glue, and industrial applications, are highly allergenic and exhibit significantly lower stability and performance at body temperatures, when compared with vertebrate collagens.\textsuperscript{41} In addition, recent findings have shown that fish also harbor prions.\textsuperscript{42}

The expanding need for collagen and its byproducts in industrial applications has spurred development of high-throughput extraction and recombinant synthesis and purification techniques, designed to provide pure material resembling its natural form. However, the complex biosynthesis of collagen, involving a relatively large number of enzymes regulating its expression and maturation, imposes considerable demands on recombinant expression systems. Mammalian cell lines have been proven effective in induction of procollagen expression and secretion, yet require mass culturing volumes, costly nutrient supplementation, and extensive time-to-product periods,\textsuperscript{43} while under constant threat of sample contamination by host pathogens. High yields of collagen are secreted in the milk of transgenic animals bearing mammary gland-targeting genomic collagen- and P4H-encoding inserts. Yet, the homotrimeric product fails to mimic the native Type I collagen heterotrimer and the hypohydroxylated state of the derived collagen resulted in relatively low thermal stability.\textsuperscript{44}

Bacterial, silkworm,\textsuperscript{45} and yeast expression systems are less expensive, when compared to mammalian cell culture systems, and appropriate for mass production of certain proteins, yet often lack enzymes and co-factors necessary for proline and lysine hydroxylation and lack of disulfide bridge formation, rendering them unfit for expression of mature and functional collagen. While coexpression of collagen and mammalian-derived prolyl-hydroxylase in insect cells enabled the formation of stable, hydroxylated collagen,\textsuperscript{46} the ability to effect other critical post-translational modifications is still required to ensure a functional mimic of natural collagen. Similar coexpression systems in yeast\textsuperscript{47,48} yielded stable, triple helical collagen, yet, yeast expression systems require relatively expensive stainless steel fermentation and support facilities.

The advent of plant-made pharmaceutics involving genetic manipulations programming plants to express molecules of therapeutic value, has introduced a feasible alternative to conventional, fermentation-based expression models. Plant engineering offers cost-effective, safe, scalable, and easily scalable protein yields harvestable after culture periods significantly shorter than in other expression systems.\textsuperscript{49} The protein synthesis pathways highly conserved between plant and eukaryote systems, along with the absence of human and animal pathogens in plants add to the attractiveness of this system. Another fundamental advantage lies in the easily harvestable leaves, tubers, and seeds that feature extended shelf lives and can serve as a simple means of storing recombinant proteins before extraction. Expression of recombinant proteins in plants can be further enhanced via sorting and targeting recombinant expression to subcellular compartments, including the nucleus, cytosol, chloroplast, apoplast, ER (reviewed in Fischer and Emans\textsuperscript{49}), and plastids.\textsuperscript{50}

The versatility of plants as bioreactors is becoming increasingly apparent when considering the wide range of plant types engineered to produce a diversity of products, from small proteins\textsuperscript{51} to complex and large antibodies.\textsuperscript{52} Production of edible recombinant proteins and vaccines have been proposed and executed in potatoes,\textsuperscript{53} lettuce,\textsuperscript{54} tomatoes,\textsuperscript{55} and alfalfa.\textsuperscript{56} Cultivated plant suspension cells allow similar protein expression under tightly regulated conditions (reviewed in Fischer et al.\textsuperscript{57}). As a non-food crop with a large leaf mass and prematurity stage harvesting, exploitation of tobacco plants as biofactories alleviate both concerns of food and feed supply contamination and of gene flow.

While the plant-based expression platform combines numerous advantages over other systems, lack of or inefficient enzymatic support often leads to plant-derived recombinant molecules void of modifications critical to their half-life and activity.\textsuperscript{58} In the case of recombinant collagen expression, plant-derived P4H can generate hydroxyproline-containing proteins, but exhibits relatively loose substrate sequence specificity in comparison to mammalian P4H.\textsuperscript{58,59} In addition, amino acid analysis of tobacco-
expressed human collagen demonstrated hydroxylysine content to be less than 2% of that found in bovine collagen, suggesting that endogenic plant LH fails to sufficiently hydroxylate collagen lysines. Hydroxylated type III collagen was successfully expressed in tobacco cells transformed with cDNAs encoding full-length collagen and human P4H \( \alpha \) and \( \beta \), but its hydroxyproline content was 16%–25% lower than human tissue-derived collagen. Moreover, no details of the hydroxylysine and O-glycosylation content are provided. Plant-specific N-glycosylation, which differs from mammalian glycans in their core residues and in the full absence of galactose and sialic acid units, presents a major limitation on plant-based expression of mammalian proteins. It is likely that plant glycans induce immunogenic responses, and that these non-native glycosylation profiles can affect the functionality of recombinant protein.

Several attempts have been made to modify the plant N-glycan biosynthesis machinery. Stable introduction of the human \( \beta_1,4 \)-galactosyltransferase to tobacco plants, then crossed with plants engineered to express the Mgr-48 mouse monoclonal IgG1, yielded galactosylated N-glycans at ratios similar to those achieved in mammalian expression systems. A strategy devised to knock out the plant \( \beta_1,2 \)-xylosyltransferase and \( \alpha_1,3 \)-fucosyltransferase, responsible for the immunogenicity of xylose- and fucose-containing glycoproteins, provides a platform in which \( \beta_1,4 \)-galactosyltransferase can then be overexpressed to act on recombinant proteins in a noncompetitive manner.

Integration of a matrix of distinct plant features toward development of a high output system for expression of both proline and lysine hydroxylated, and O-glycosylated heterotrimeric, recombinant human procollagen type I has been reported. Through a series of crossbreedings, two human genes encoding recombinant heterotrimeric collagen type I (rhCOL1) were successfully coexpressed in tobacco plant vacuoles with the human P4H-\( \alpha \) P4H-\( \beta \) and LH3 enzymes. Plants coexpressing all five vacuole-targeted proteins generated intact procollagen at yields of 1 g/kg dry tobacco leaves (Fig. 1). The isolated protein contained heterotrimers only, as demonstrated by collagenase treatment, to which homotrimeric collagen is resistant. In addition, the hydroxylated proline and lysine content was within the range of that found in human collagen (7%–10% and 0.7%–1%, respectively). As in natural processes, underhydroxylated collagen molecules were degraded and negatively selected in the course of the enzymatic processing of procollagen to atelocollagen. O-glycosylation was low and was identical to natural levels, as indicated by mass spectrometry. The acidic and easily extractable plant vacuole provided an ideal intracellular subcompartment for procollagen expression. The closed environment protected maturing collagen molecules from undesirable N-glycosylations, which sequentially occur on the cytosolic and luminal faces of the ER. Plant-extracted rhCOL1 formed thermally stable triple helical structures and demonstrated fibril-forming capacities (close to 100%) and biofunctionality similar to human tissue-derived collagen supporting attachment and expansion of adult various primary human cells normally involved in tissue repair processes.

**FIG. 1.** Production of recombinant human collagen in tobacco plants. (A) Engineered tobacco plants are easily grown in large masses under regulated greenhouse conditions. (B) The tobacco plant’s high leaf mass renders it ideal for collection for large-scale protein extraction. (C) Tobacco plants were engineered to express five different genes that include two human collagen type I subunits, along with three human enzymes critical for post-translational modification of procollagen type I. All five proteins were specifically designed to be diverted to the plant vacuole. (D) Procollagen type I is extracted from the transgenic tobacco leaves. (E) The procollagen is purified from the protein extract and processed. (F) Pure human recombinant collagen is produced. Color images available online at www.liebertpub.com/tea
Closing Remarks

Deployment of the highly versatile plant-based biofactory can be leveraged toward mass, rapid, and low-cost production of a wide variety of recombinant proteins, high-value crops and pharmaceutics. As in the case of collagen, proper planning can bypass plant-related limitations, to yield products structurally and functionally identical to their native counterparts.

Disclosure Statement

F.G. is employed by CollPlant, Ltd. O.S. is the founder and chief scientific officer of CollPlant, Ltd.

References


Address correspondence to:
Oded Shoseyov, PhD
The Robert H. Smith Faculty of Agriculture
Food and Environment
The Robert H. Smith Institute of Plant Science and Genetics
The Hebrew University of Jerusalem
P.O.B. 12
Rehovot 76100
Israel
E-mail: shoseyov@agri.huji.ac.il

Received: June 4, 2012
Accepted: December 11, 2012
Online Publication Date: February 18, 2013