Production of Human Type I Collagen in Yeast Reveals Unexpected New Insights into the Molecular Assembly of Collagen Trimers

Received for publication, February 21, 2001

Published, JBC Papers in Press, March 14, 2001, DOI 10.1074/jbc.M101613200

David R. Olsen‡, Scott D. Leigh‡, Robert Chang‡, Hugh McMullin‡, Winson Ong‡, Ernest Tai‡,
George Chisholm§, David E. Birk¶, Richard A. Berg‡, Ronald A. Hitzeman§, and P. David Toman¶

From ‡Cohesion Technologies Inc., Palo Alto, California 94303, §Genotypes, Inc., South San Francisco, California 94080, and ¶Thomas Jefferson University, Jefferson Medical School, Philadelphia, Pennsylvania 19107

Substantial evidence supports the role of the procollagen C-propeptide in the initial association of procollagen polypeptides and for triple helix formation. To evaluate the role of the propeptide domains on triple helix formation, human recombinant type I procollagen, pN-collagen (procollagen without the N-propeptide), pC-collagen (procollagen without the N-propeptide), and collagen (minus both propeptide domains) heterotrimers were expressed in Saccharomyces cerevisiae. Deletion of the N- or C-propeptide, or both propeptide domains, from both proo-chains resulted in correctly aligned triple helical type I collagen. Protease digestion assays demonstrated folding of the triple helix in the absence of the N- and C-propeptides from both proo-chains. This result suggests that sequences required for folding of the triple helix are located in the helical/telopeptide domains of the collagen molecule. Using a strain that does not contain prolyl hydroxylase, the same folding mechanism was shown to be operative in the absence of prolyl hydroxylase. Normal collagen fibrils were generated showing the characteristic banding pattern using this recombinant collagen. This system offers new opportunities for the study of collagen expression and maturation.

Collagen is the single most abundant protein found in animals. In the human body, it is expressed in most tissues and plays a structural, as well as a signaling, role in the development, maintenance, and repair of tissues and organs. 20 different collagen types are coded by more than 30 genes. Assembly of trimeric collagen intracellularly and formation of collagen fibers in the extracellular matrix is the result of a complex multistep process (1, 2). Within the endoplasmic reticulum, the individual procollagen polypeptides undergo several co- and post-translational modifications, including hydroxylation of specific proline and lysyl residues, selection and alignment of three procollagen polypeptides, and disulfide bond formation among the C-propeptides. Experimental evidence suggests folding of the triple helix begins at the C terminus and propagates toward the N terminus. Prior to triple helix formation, prolyl hydroxylase converts proline in the Y position of GXY triplets to hydroxyproline. Hydrogen bonding between the α-chains of the triple helix increases the denaturation temperature of the molecule, preventing it from unfolding at the animal’s body temperature (3). Triple helical procollagen is secreted from the cell, and the N- and C-terminal propeptides are removed by specific N- and C-proteinases. The resulting collagen monomers, consisting of triple helical and telopeptide regions, undergo a self-assembly process to generate collagen fibril intermediates and then mature collagen fibers. These fibers are further stabilized by covalent cross-links within the triple helix and telopeptide regions, providing strength and support to the surrounding tissue.

Collagen biosynthesis has been studied extensively for many years, but the mechanism by which three procollagen polypeptides initially associate and come into correct registration is not well understood. Early work studying procollagen assembly suggested that the C-propeptide plays an undefined but critical role in assembly of a triple helical molecule (4–6). More recently, the procollagen C-propeptide has been implicated in selection of appropriate procollagen polypeptides and correct registration of the triple helix (7). The C-propeptide of type III procollagen is reported to direct the initial association of polypeptides and is required for helix formation (8). Expression of α1(III)-α2(II) procollagen C-propeptide chimeras led to the identification of a discontinuous sequence of 15 amino acids, GNELPDEDVLDV...SSR, that directed procollagen self-association (9). Other reports implicate the last 10 amino acids of the α2(I) procollagen C-propeptide, as well as intramolecular disulfide bond formation, as prerequisites for assembly of the type I procollagen heterotrimer (10, 11).

Our studies further address the role of the procollagen polypeptides, especially the role of the C-propeptide of type I procollagen, in folding of the triple helix. Different yeast strains have been shown to express and correctly fold human procollagens (12–14). The Saccharomyces cerevisiae expression system allowed us to evaluate the minimal requirements for collagen synthesis and folding, because these cells most likely lack any specialized genes for collagen biosynthesis in fibroblastic cells. We show that type I collagen heterotrimers expressed in yeast were correctly assembled into triple helical collagen monomers in the absence of the N- and C-propeptide domains from the proc1(I) and proc2(I) collagen polypeptides. Furthermore, we show that this same folding mechanism is operative in yeast in the absence of prolyl hydroxylase. The recombinant collagen forms normal fibrils in vitro.

EXPERIMENTAL PROCEDURES

Plasmid Constructions—Plasmid pGET323 (similar to pGET462 shown in Fig. 1A minus the preproHSA-α2(I) procollagen-PGKt; described in Ref. 12) was used to create plasmids encoding α1(I) procollagen lacking the N-propeptide, C-propeptide, or both propeptides. pGET323 contains the yeast TRP1 gene, 2 micron origin, Escherichia coli β-lactamase gene, and CoE1 origin (15), as well as the yeast GAL1–10 dual promoter (16), driving the production of preproHSA-α1(I) procollagen. A plasmid encoding type I procollagen homotrimer

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† To whom correspondence should be addressed: Cohesion Technologies Inc., 2500 Faber Place, Palo Alto, CA 94303. Tel.: 650-320-5595; Fax: 650-320-5511; E-mail: dtoman@cson.com.
was created using pGET323 as a template for PCR with the following primers: 5'-AGCGTGCAAGCTTTCTTTTCTGGAAG-3' (sense) and 5'-TTGGAAGCTTGGGGACCAGGTGCA-3' (antisense). This and all PCRs used the following program: 94 °C, 1 min; 60 °C, 1 min; 72 °C, 2 min; 35 cycles. All DNAs generated by PCR were sequenced to confirm the desired sequence. The PCR DNA was purified and digested with SalI-ApaLI. pGET323 was digested with SalI and DraIII, treated with calf intestinal phosphatase, and the 8.1-kb fragment was isolated. The 1.1-kb DraIII-ApaLI fragment from pGET323 was isolated and three fragments were ligated to create pDO243858, encoding α1(I) p-collagen with the N-propeptide sequence removed and the preproHSA secretion signal fused directly to the C-terminal telopeptide. pDO243858 was further modified by introduction of the N-telopeptide. pDO243858 was ligated to create pDO248015, creating plasmid pDO243858. This was accomplished using an intermediate plasmid that contained the 3' half of the α1(I) procollagen cDNA from the BanHI site at nucleotide 2803, relative to the ATG initiator codon, to the SalI site in the 3' untranslated region. The plasmid was digested with EcoRI and BspEI, and the following sequence was cloned overlapping oligonucleotides: 5'-GCTGTGGTGTACATCCCTCCTCCTCCCCAGGCCCACC-ACCTCAAAGAGCCGTCGTCCAGCAGACGGAGAGATGATGCCAATGTGGTTCGTGACCGTGACCTCGAGGTCGACACC-ACCTCAGGCTACGAGGCAGCCAGACCATGATGAAATCGTCTTCAT-CTCTTCGCTCCTGTTTTCGCTGACATGAGCAGCTGACGAG-3' (sense) and the corresponding antisense oligonucleotide to create a double stranded oligonucleotide that introduces the corresponding antisense oligonucleotide to create a double stranded DNA. The PCR DNA was purified and digested with SauI and NsiI, and the 8.141-kb fragment was ligated to the 5.494-kb NsiI-AgeI fragment from pDO243873, creating pDO243865. The 5.494-kb NsiI-AgeI fragment from pDO248050 was ligated to the 6.922-kb NsiI-SacII fragment from pDO248040, creating plasmid pDO243869. A plasmid encoding type I p-collagen heterotrimer, lacking the N-propeptide sequences from both α1(I) and α2(I) procollagen polypeptides, was constructed by ligating the 5.494-kb NsiI-AgeI fragment from pDO243858, and the 1.213-kb SacII-AgeI fragment from pDO243866, and the 8.141-kb fragment from pDO243873. Plasmids pDO243869 and pDO243865 were transformed into yeast GY5382 to create strains CTY89, -87, -90, and -59, respectively, and grown at 30 °C. Plasmid pDO248055 was also transformed into yeast GY5316 to create strain CYT38 and was grown at 20 °C. Yeast strains were transformed by electroporation (17) and grown in 2% glucose, 0.67% yeast nitrogen base with ammonium sulfate, and 0.5% casamino acids. Induction of protein expression was performed in medium similar to that described above except that a mixture of 0.5% glucose and 1% galactose was used as the carbon source, and the medium was buffered with 1% sodium citrate, pH 6.5. GY5196 strains contained 2% galactose as the sole carbon source during induction, and cultures were supplemented with 34 μg/ml uracil.

**Fermentation and Purification**—Recombinant collagen was purified from yeast cells following fermentation at the 10-liter scale (Biostat C; B. Braun, Goettingen, Germany). Fermentation conditions were maintained at 30 °C (strain GY5382) or 20 °C (strain GY196) (both strains as described in Ref. 12), pH 6.5, and aeration at 1 air volume/liquid volume/min. Dissolved oxygen was controlled at 20% by automatically adjusting the impeller speed. The fermentation was started batchwise with medium composed of 5 g/liter casamino acids, 40 g/liter glucose, 0.5 g/liter galactose, and 13.4 g/liter yeast nitrogen base (YNB; Difco). Galactose levels were monitored throughout the fermentation and maintained at 0.5 g/liter. When the glucose was depleted from the medium, the culture was fed to achieve a final concentration of 20 g/liter glucose with 6.7 g/liter YNB added to supplement the growth medium. When the glucose from the first feed was depleted, a second feed of 20 g/liter glucose and 6.7 g/liter YNB was added to the fermentor. The yeast were harvested when cell growth ceased and were frozen at -70 °C. Frozen cell paste was resuspended in 0.2 M Tris-HCl (pH 7.4, 0.4 M NaCl and passed through a DYNOMILL (Type KDIA; Bachofen AG, Basel, Switzerland) to rupture the cells. Cell debris was removed by centrifugation at 10,000 × g for 30 min at 4 °C, and the supernatant, containing total soluble protein, was collected and delipidated using Celite 512. After delipidation, collagen was recovered from the soluble fraction by NaCl precipitation at neutral pH values as described (18). The protein was dialyzed against 0.1 M NaPi buffer, pH 7.4, 0.4 M NaCl and dialyzed against 0.1 M NaPi buffer, pH 4.5 (Buffer A), at 4 °C. The precipitate that formed during dialysis was removed by centrifugation, and the supernatant containing the collagen was further purified by cation exchange chromatography. A column of SP-Sepharose (5 × 18 cm) was equilibrated in Buffer A at 4 °C, and the sample was applied to the column at a flow rate of 5 ml/min. Bound proteins were eluted with a linear NaCl gradient (0–0.5 M NaCl, 20 column volumes). Fractions containing purified collagen were pooled, and the protein was recovered by precipitation at an acidic pH value with NaCl (18). The precipitate was solubilized in 10 mM HCl at a concentration of ~3 mg/ml.

**Protease Digestion Assays**—Rhodocill I samples were assayed for helix formation and thermal stability by pepstatin or trypsin/chymotrypsin digestion (19). Samples were digested with 150 μg/ml pepstatin for 30 min at 10, 20, 30, 40, and 45 °C. The reaction was terminated by neutralization of the sample with 1 NaOH. Trypsin/chymotrypsin digestion were performed as described previously (19), with minor modifications. Prior to addition of the proteases, samples were preincubated at the desired temperatures (20, 25, 30, 35, and 45 °C) for 15 min. The proteases were added and incubated for an additional 2 min. Partially purified type I collagen was obtained from serum-free conditioned medium of human skin fibroblast cultures by 20% ammonium sulfate precipitation. Precollagen was treated with proteases as a control in these assays. Reaction products were analyzed by SDS-PAGE using 6% gels, and proteins were visualized by staining with Gelcode Blue (Pierce, Rockford, IL). Scanning densitometry was done using the Bio-Rad Gel Doc 1000 gel scanner.
**RESULTS**

Expression of Collagen with Deleted Propeptide Regions—A previously described heterologous yeast expression system (12) was used to directly examine the role of the propeptides on folding of the type I procollagen triple helix. A plasmid was constructed that contained α1(I) and α2(I) procollagen cDNAs fused to the human serum albumin signal sequence under the control of the GAL1–10 promoter. Additional constructs were made that did not contain the N-propeptide, C-propeptide, or both propeptides (Fig. 1). This series of plasmids were transformed into a *Saccharomyces cerevisiae* strain that was previously engineered to express functional chicken prolyl hydroxylase. Collagen expression levels were measured with a quantitative immunoassay that specifically detects native type I collagen heterotrimers but is unable to detect denatured type I collagen or type I collagen homotrimers (12). Expression of α1(I) and α2(I) procollagen heterotrimer lacking the N-propeptide (pC-collagen) resulted in detection of a 3.9-fold increase in expression relative to full-length procollagen (Table I). Surprisingly, a construct expressing α1(I) and α2(I) procollagen polypeptides lacking the C-propeptide (pN-collagen) was detected by the immunoassay; furthermore, expression was 5.9-fold higher than procollagen in the same strain. Because the assay detected pN-collagen, it strongly suggested that the pN-collagen polypeptides had folded into a triple helix. The generation of an expression plasmid that codes only for the telopeptide and C-propeptide, whereas the triple helix is resistant to these proteases. Additionally, collagen α-chains that are not correctly folded into a triple helix are susceptible to proteolysis (18). Protease digestion assays were performed to probe the helical conformation of the yeast-derived rhc collagen I and demonstrated that the chains were correctly registered and that the rhc collagen I was of the expected molecular weight. Purified

**TABLE I**

| Strain | Proteins encoded | Collagen expression levels (µg collagen/mg protein) |
|--------|------------------|--------------------------------------------------|
| CYT89  | α1(I)/α2(I) procollagen | 1.64                                             |
| CYT90  | α1(I)/α2(I) pC-collagen | 6.45                                             |
| CYT87  | α1(I)/α2(I) pN-collagen | 9.7                                              |
| CYT59  | α1(I)/α2(I) rhc collagen | 29.8                                             |
Pepsin digestion

Trypsin/chymotrypsin digestion

yeast-produced non-hydroxylated rhcollagen I

yeast-produced hydroxylated rhcollagen I

procollagen (control)

**Fig. 2. SDS-PAGE gels of protease-treated yeast-produced rhcollagen I at different temperatures.** Non-hydroxylated rhcollagen I (panel A), hydroxylated rhcollagen I (panel B), or procollagen (panel C) were digested with 150 μg/ml pepsin at 10, 20, 30, 40, and 45 °C (lanes 4–8) or were not digested (lane 3). Non-hydroxylated rhcollagen I (panel D), hydroxylated rhcollagen I (panel E), and procollagen (panel F) were digested with trypsin/chymotrypsin at 20, 25, 30, 35, 40, and 45 °C (lanes 4–9) or were not digested (U; lane 3). Lane 1 contains a molecular weight marker (M) (broad range marker; Bio-Rad), and lane 2 contains a type I collagen standard (S). Digests were fractionated by SDS-PAGE using 6% gels, and proteins were visualized by staining with GelCode Blue.

non-hydroxylated and hydroxylated rhcollagen I and procollagen were treated with pepsin at different temperatures (Fig. 2, A–C). Non-hydroxylated rhcollagen I was resistant to pepsin digestion at 10 and 20 °C, whereas hydroxylated rhcollagen I was resistant to pepsin treatment up to 30 °C. The non-hydroxylated collagen polypeptides migrated slightly faster in SDS-PAGE gels than hydroxylated polypeptides (Fig. 2A, compare lanes 2 and 3), a finding consistent with previous reports of altered α-chain migration in gels because of the lack of hydroxylation (23). The fibroblast-derived procollagen control was converted to collagen and was pepsin-resistant up to 30 °C but was degraded at 40 °C. Resistance to pepsin alone should be evaluated with caution when judging the formation of a collagen triple helix (24), with recommendation to use a combination of trypsin and chymotrypsin to probe the integrity and proper alignment of the triple helix. The rhcollagen I and fibroblast-derived procollagen were also treated with a mixture of trypsin and chymotrypsin, and the results of digestions with pepsin were confirmed with this mixture of proteases (Fig. 2, D–F). With the exception of the slight increase in mobility of the non-hydroxylated collagen, the size of the protease-resistant bands in both sets of digestions were the same for native and recombinant collagen, indicating that recombinant non-hydroxylated and hydroxylated rhcollagen I molecules have formed a correctly aligned, full-length triple helix. Furthermore, these results suggest that folding of the triple helix in yeast can occur not only in the absence of both N- and C-propeptides but also without prolyl hydroxylase.

The ratio of pepsin-resistant α1 and α2 collagen polypeptides from hydroxylated rhcollagen I was not 2:1 as expected for type I collagen heterotrimer but was estimated to be 5:1 by scanning densitometry of SDS-PAGE gels. Our interpretation of this finding is that both heterotrimeric and homotrimeric (α1)3 rhcollagen I molecules are present in approximately equal amounts. The presence of homotrimeric type I collagen has been observed in tissue, certain fibroblasts producing type I collagen, and in other recombinant expression systems (25–27). A further test of the integrity and alignment of the triple helix of rhcollagen I was cleavage using mammalian collagenase. Mammalian collagenase makes one specific cleavage in all three polypeptides in the triple helix at position 775–776. In addition to a specific primary sequence Gly-(Ile/Leu)-(Ala/Leu), the local structure and fold of the helix is also critical for collagenase cleavage (28, 29). If the rhcollagen I triple helix were out of alignment by as little as 3 amino acids, collagenase cleavage would not occur (30). Digestion of native type I collagen and hydroxylated rhcollagen I with mammalian collagenase resulted in the generation of typical collagenase reaction products (Fig. 3), indicating both rhcollagen I hetero- and homotrimmers were cleaved by collagenase. The larger fragments of both α1 and α2(1) collagen (TC75; Lane 1) are visible on the gel, as well as the smaller fragments (TC76) from native type I collagen. Consistent with the lower ratio of α2 versus α1 rhcollagen I polypeptides, the proteolytic TC76 fragment of α2 rhcollagen I was not visible on this gel. These results show that the polypeptides of rhcollagen I are properly aligned.

Hydroxylated and non-hydroxylated rhcollagen I was compared with human placental type I collagen by CD spectroscopy over a range of temperatures. The T_m values determined by CD analysis were 30, 24, and 40 °C for hydroxylated and non-hydroxylated rhcollagen I and native type I collagen, respectively (Fig. 4). These results suggest the rhcollagen I produced by the strain containing the prolly hydroxylase genes was partially hydroxylated. This finding was confirmed by amino acid analysis showing the rhcollagen I had 49 ± 2% (mean ± standard deviation) of the hydroxyproline expected for collagen isolated from tissue and less than procollagen expressed in the same strain (12). Hydroxylated lysine residues were not detected.
measurements are shown as equilibrated at each temperature and then scanned. Temperature circles (CYT38; recombinant collagen).

...with pepsin and trypsin/chymotrypsin, we demonstrated resistance of the protease digestion. Proteolytic enzymes digest single polypeptide 4–11). close proximity to each other to form a folding nucleus (1–2, 21, 22). The fibrils showed a periodicity of 67 nm, comparable with in vivo fibrillogenesis experiments using type I collagen and with fibers found in most tissues. The melting temperature of fibrils made from rhcollagen I was measured at 48 °C using differential scanning calorimetry, comparable with native bovine collagen isolated from hide (51–52 °C; data not shown).

**DISCUSSION**

Previous studies showed that triple helical type I procollagen could be generated in transfected mammalian cells with α1 and α2(I) procollagen polypeptides lacking the N-propeptide (31). To further study the role of the propeptide domains in collagen triple helix assembly, engineered α1(I) and α2(I) procollagen genes were generated that encode procollagen lacking either the N- or C-terminal propeptide regions or both propeptide regions. Using an S. cerevisiae heterologous expression system, type I collagen polypeptides were assembled into correctly aligned triple helix without the presence of the C-propeptide. These results indicated that only the triple helix and telopeptide regions of the molecule are sufficient for assembly of a triple helix. These results were completely unexpected based on all of the reports describing the essential role of the C-propeptide in pros-chain selection and registration to bring the sequences at the C terminus of the triple helical region in close proximity to each other to form a folding nucleus (1–2, 4–11).

The triple helix of rhcollagen I was further examined using protease digestion. Proteolytic enzymes digest single polypeptide chains and incorrectly folded collagen molecules. In two separate protease digestion experiments, using either pepsin or trypsin and chymotrypsin, we demonstrated resistance of the triple helix of rhcollagen I to these proteases, indicating that correct folding of rhcollagen I had occurred. Additionally, correctly folded full-length rhcollagen I was produced by yeast cells that were not engineered to express prolyl hydroxylase. Thus, neither proline hydroxylation nor the prolyl hydroxylase enzyme is essential for polypeptide chain association and folding of the triple helix under our experimental conditions.

We questioned whether the pepsin and trypsin/chymotrypsin-resistant rhcollagen I, which appears full-length by SDS-PAGE, may contain α1 and α2 rhcollagen I chains that are out of register by only a few residues. If this was the case, proteolytic removal of a few N- or C-terminal amino acids in a non-helical configuration because of α-chain misalignment would not be detected by SDS-PAGE. To distinguish between correctly aligned and slightly misaligned α-chains, we used mammalian collagenase digestion as a probe. Collagenase cleaves collagen at one specific locus, the Gly-Ile bond at position 775–776. In addition to this primary sequence requirement for cleavage, other parameters affect cleavage (28). Type I, II, and III collagen contain the partial sequence Gly-(Ile/Leu)-(Ala/Leu) at 27 other locations throughout their helical domains. All of these other loci are not cleaved by collagenase. The reasons for this strict specificity is not completely understood but suggests that extended sequences around the cleavage site or a specific folded structure are required for activity. It has been proposed that the helix is tightly folded up to and including the Gly residue at the cleavage site. Because the four triplets following Gly-775 are imino acid deficient, a loosely folded helical structure is thought to exist at the C-terminal side of the cleavage site (29). The border between the tight and loose forms of the helix may act to present the cleavage site more efficiently to collagenase. We reasoned that if our collagen chains were even slightly out of alignment, these additional structural requirements of collagenase would be altered or lost. Furthermore, mutagenesis experiments have shown that alteration of the primary sequence of α1 inhibits cleavage, and the presence of non-cleavable α1-chains in a heterotrimer with wild-type α2-chains results in inhibition of α2-chain cleavage (32). Cleavage of the α1- and α2-chains of rhcollagen I was complete, indicating both hetero- and homotrimmers were cleaved by collagenase. This finding suggests that the α-chains of rhcollagen I are properly aligned.

Fibrillogenesis data attest to the integrity and quality of...
New Insights into Molecular Assembly of Collagen Trimers

this recombinant collagen. Electron micrographic analysis of fibers formed with rhcollagen I shows the typical banding pattern of fibers seen in tissues. Telopeptides play an important role in collagen fibrillogenesis as their removal alters both the kinetics of fiber formation and fiber morphology including loss of diameter uniformity and unidirectional packing (33). The presence of the telopeptide domains adds additional physical properties to rhcollagen I that may be lacking in protease-treated collagen. Even though the rhcollagen I monomer has a hydroxyproline level at ~50% of the level of fully hydroxylated collagen, these molecules produce a fibril with similar thermal stability to tissue-derived collagen.

Deletion of the N-propeptide and C-propeptide and the combination of the two deletions resulted in dramatic increases in collagen levels. One possible explanation of this phenomena is that synthesis of proα-chains represents the rate-limiting step in collagen production in yeast. Removal of ~45 Kd of protein (~15 Kd from the N-propeptide and ~30 Kd from the C-propeptide) from each α-chain may reduce a significant biosynthetic burden, resulting in higher expression levels. Additional post-translational modifications occur in the propeptides, including N-linked glycosylation of the C-propeptide and the formation of several disulfide bonds in both propeptides. The elimination of these processing steps could also enhance expression.

A second explanation relates to the role of the propeptides in folding. The C-propeptide has been thought to play a key role in pro-chain selection and registration (4–7). Additionally, the C-propeptide may also affect the rate of folding, acting to limit helix formation, thus allowing prolyl hydroxylase to bind to the unfolded proα-chains and completely hydroxylate proline residues in the Y position of GXY triplets. If the C-propeptides were removed and the chains were allowed to fold faster, expression levels would be expected to increase. An additional consequence of this increased folding rate would be decreased hydroxylation levels, because helix formation limits hydroxylation. Our data indicate the rhcollagen I is underhydroxylated, supporting this hypothesis. Furthermore, when the same yeast background was used to express procollagen, hydroxylation levels were 33% higher (12). The lower levels of hydroxylation seen in the rhcollagen I strain may be overcome by increasing the levels of prolyl hydroxylase.

In this study we demonstrated selection and folding of type I rhcollagen I subunits into (α1)2α2 heterotrimers and (α1)3 homotrimers in yeast in the absence of propeptide sequences. We have engineered type I collagen genes that encode the N- and C-telopeptides with the entire triple helical domain and show that these sequences are sufficient for assembly of a triple helix in S. cerevisiae. Other fibrillar collagens (types II-III, V, and XI) have a similar structure and thus would be expected to fold into triple helices without the propeptide regions in an analogous system. Additionally, manipulating the collagen molecule in this manner has allowed us to evaluate folding mechanisms separate from hydroxylation and other roles played by prolyl hydroxylase.

Yeast has been used as a model system to study the expression, trafficking, and assembly of eukaryotic proteins, and this organism is now being utilized for studying the mechanisms of collagen synthesis and folding (12–14). The minimal requirements for collagen synthesis and folding were evaluated, because yeast do not express collagen and are unlikely to possess the specialized genes required for collagen biosynthesis in other cells (such as fibroblasts). Our studies in this system have provided new insights into the minimal structural requirements for collagen expression and folding. We are in the process of testing additional collagen types using this system and also are testing mammalian cells for the functional requirements of structural domains to form a triple helical collagen. The ability to dissect the process of collagen assembly using this heterologous expression system exemplifies its utility for furthering the understanding of the key components of collagen triple helix assembly and collagen biology.

REFERENCES

1. Haralson, M. A., and Hassell, J. R. (1995) in Extracellular Matrix: A Practical Approach (Haralson, M. A., and Hassell, J. R., eds) pp. 1–30, IRL Press, New York.
2. Kadler, K. (1994) Protein Profile 1, 519–638
3. Berg, R. A., and Prockop, D. J. (1973) Biochem. Biophys. Res. Commun. 52, 115–120
4. Rosenblum, J., Endo, R., and Harash, M. (1976) J. Biol. Chem. 251, 2070–2076
5. Schufeld, D. J., Uitto, J., and Prockop, D. J. (1974) Biochem. 13, 1801–1806
6. Uitto, V., Uitto, J., and Prockop, D. J. (1981) Arch. Biochem. Biophys. 210, 445–454
7. McLaughlin, S. H., and Bulleid, N. J. (1998) Matrix Biol. 16, 369–377
8. Bulleid, N. J., Dalley, J. A., and Lee, J. F. (1997) EMBO J. 16, 6684–6701
9. Lees, J. F., Sasab, M., and Bulleid, N. J. (1997) EMBO J. 16, 908–916
10. Lim, A.-L., Doyle, S. A., Balian, G., and Smith, B. D. (1998) J. Cell. Biochem. 71, 216–232
11. Doyle, S. A., and Smith, B. D. (1998) J. Cell. Biochem. 71, 233–242
12. Toman, P. D., Chiasholm, G., McMillin, H., Gier, L. M., Olsen, D. R., Kovach, R. J., Leigh, S. D., Fong, B. E., Chang, R., Daniels, G. A., Berg, R. A., and Hitzeman, R. A. (2000) J. Biol. Chem. 275, 23303–23309
13. Vaughn, P. R., Ganalis, M., Richards, K. M., Tebb, T. A., Ramshaw, J. A. M., and Werkmeister, J. A. (1998) DNA Cell Biol. 17, 511–518
14. Vuorela, A., Myllyharju, J., Nissi, R., Pihlajaniemi, T., and Kivirikko, K. I. (1997) EMBO J. 16, 6702–6712
15. Chen, C., Oppermann, H., and Hitzeman, R. (1984) Nucleic Acids Res. 12, 8891–8970
16. Johnston, M., and Davis, R. (1984) Mol. Cell. Biol. 4, 1440–1448
17. Becker, D. M., and Guarente, L. (1991) Methods Enzymol. 194, 182–187
18. Miller, E. J., and Rhodes, R. K. (1982) Methods Enzymol. 82, 33–64
19. Bruckner, P., and Prockop, D. J. (1981) Anal. Biochem. 110, 360–368
20. McPherson, J. M., Wallace, D. G., Sawamura, S. J., Conti, A., Condel, R. A., Wade, S., and Piez, K. A. (1985) Collagen Relat. Res. 5, 119–135
21. Birk, D. E., Fitch, M. J., Babinz, J. P., Doane, K. J., and Linsenmayer, T. F. (1990) J. Cell Sci. 95, 649–657
22. Birk, D. E., Nurminskaya, M. V., and Zychiel, E. I. (1995) Dev. Dyn. 202, 229–243
23. Kato, K., Prockop, D. J., and Berg, R. A. (1979) J. Biol. Chem. 254, 2234–2243
24. Bulleid, N. J., Wilson, R. R., and Lad, U. (1995) Matrix Biol. 14, 233–236
25. Uitto, J. (1979) Arch. Biochem. Biophys. 192, 371–379
26. Lee, S.-T., Smith, B. D., and Greenspan, D. S. (1988) J. Biol. Chem. 263, 13414–13418
27. Myllyharju, J., Lamberg, A., Nothomb, H., Fietzek, P. P., Pihlajaniemi, T., and Kivirikko, K. I. (1997) J. Cell Biochem. 72, 2184–2180
28. Hightberger, J. J., and Corbett, C. J. (1979) J. Biol. Biochem. Res. Commun. 88, 202–208
29. Fields, G. B., and Van Wart, H. E. (1992) Matrix Suppl. 1, 68–70
30. Byrne, M. H., Wu, H., Birdhead, J. R., Goldring, M. B., Jaenisch, R., and Krane, S. M. (1993) Matrix 15, 30
31. Lee, S.-T., Lee, S., Peters, D. P., Hoffman, G. G., Stacey, A., and Greenspan, D. S. (1992) J. Biol. Chem. 267, 24316–24313
32. Wu, H., Byrne, M. H., Stacey, A., Goldring, M. B., Birdhead, J. R., Jaenisch, R., and Krane, S. M. (1990) Proc. Natl. Acad. Sci. U. S. A. 87, 5888–5892
33. Gelmans, R. A., Poppke, D. C., and Piez, K. A. (1979) J. Biol. Chem. 254, 11741–11745
34. Hitzeman, R. A., Chen, C. Y., Dowbenko, D. J., Renz, M. E., Liu, C., Pai, R., Simpson, N. J., Kohr, W. J., Singh, A., Chiasholm, V., Hamilton R., and Chang, C. N. (1990) Methods Enzymol. 185, 421–440