Synthesis of recombinant human procollagen II in a stably transfected tumour cell line (HT1080)

Andrzej FERTALA,* Aleksander L. SIERON,* Arupa GANGULY,* Shi-Wu LI,* Leena ALA-KOKKO,* Kalyan R. ANUMULA† and Darwin J. PROCKOP*‡

* Department of Biochemistry and Molecular Biology, Jefferson Institute of Molecular Medicine, Jefferson Medical college, Thomas Jefferson University, Philadelphia, PA 19107, U.S.A., and † Macromolecular Sciences and Analytical Chemistry Departments, SmithKline Beecham Pharmaceuticals, King of Prussia, PA, U.S.A.

Apparently because the biosynthetic pathways involve eight or more highly specific post-translational enzymes, it has been difficult to obtain expression of genes for fibrillar collagens in recombinant systems. Here two constructs of the human gene for procollagen II (*COL2A1*) were prepared, one with about 0.5 kb of a promoter for a procollagen I gene (*COL1A1*) and the other with about 4 kb of the promoter for the procollagen II gene. The constructs, together with a neomycin-resistant gene, were transfected into a human tumour cell line (HT1080) that synthesizes the collagen IV found in basement membranes, but does not synthesize any fibrillar collagen. About two per 100 clones resistant to the neomycin analogue G418 synthesized and secreted human procollagen II. Milligram quantities of the recombinant procollagen II were readily isolated from the cultured medium. The recombinant procollagen II had the expected amino acid sequence as defined by nucleotide sequencing of mRNA-derived cDNA and the expected amino acid composition as defined by

analysis of procollagen II that was converted into collagen II by digestion with procollagen N- and C-proteinases. Also, analysis of the carbohydrate content indicated that there was glycosylation of some of the hydroxylysine residues but no evidence of post-translational overmodification of the residues. In addition, the protein was shown to have a native conformation as assayed by a series of protease digestions. No essential differences were found between clones transfected with the *COL2A1* gene construct containing the *COL1A1* promoter and the similar construct containing the *COL2A1* promoter in terms of number of clones synthesizing recombinant procollagen II and the levels of expression. With both constructs, the expression of the *COL2A1* gene was closely related to copy number. The results demonstrated therefore that it is not essential to use a promoter for a gene normally expressed in a host cell in order to obtain gene copy-number-dependent expression of an exogenous collagen gene in stably transfected cells.

INTRODUCTION

Expression of genes for fibrillar collagens has proven to be difficult in recombinant systems, apparently because biosynthesis of the proteins requires processing by eight or more specific post-translational enzymes (Prockop and Kivirikko, 1984). In particular, post-translational modification of the nascent pro α chains of the proteins by prolyl 4-hydroxylase appears critical, as the chains cannot fold into a triple-helical conformation that is stable at 37 °C unless about 100 prolyl residues are converted into hydroxy-4-prolyl residues. Hybrid recombinant procollagens I have been synthesized in two different systems. In one system, Mov13 mouse fibroblasts that did not synthesize pro α 1(I) chains were transfected with a gene for pro α 1(I) chains, and hybrid molecules of recombinant pro α 1(I) chains and endogenous mouse pro α 2(I) were generated (Schneke et al., 1987). In similar experiments, rat fibroblasts that did not synthesize pro α 2(I) chains were transfected with a gene for pro α 2(I) chains, and hybrid molecules of procollagen I were synthesized (Lee et al., 1989).

Recently, we described a system in which the *COL2A1* gene for human procollagen II was expressed in stable transfectants of mouse NIH 3T3 cells (Ala-Kokko et al., 1991). The constructs used in the experiments contained a 5' fragment from the *COL1A1* gene that extended from –500 to +1445 bp of the gene and therefore contained part of the promoter region, the first exon and most of the first intron. The 5' fragment from the *COL1A1* gene was ligated to a 14 kb and a 12 kb *SphI* fragment which together contained most of the human *COL2A1* gene.

Stably transfected cells resistant to the neomycin analogue G418 were shown to synthesize human procollagen II, apparently because the human *COL1A1* promoter in the construct was active in the NIH 3T3 cells which synthesize mouse procollagen I but not procollagen II or any other cartilage-specific protein.

Although the recently reported system for expression of the human *COL2A1* gene in NIH 3T3 cells provided a source of recombinant human procollagen II (Ala-Kokko et al., 1991), purification of the human procollagen II from mouse procollagens I and III synthesized by the same cells has proven difficult (A. Fertala, A. L. Sieron and D. J. Prockop, unpublished work). Therefore we have examined the possibility of expressing the human *COL2A1* gene in HT1080 cells, a human tumour cell line that synthesizes the collagen IV found in basement membranes but not any fibrillar collagen (Tryggvason et al., 1980; Pihlajaniemi et al., 1981). The results demonstrate that considerable amounts of human procollagen II can be generated by the system, and that the recombinant protein can be effectively purified from the medium.

MATERIALS AND METHODS

Design and preparation of DNA constructs

Two cosmid constructs were prepared which contained most or all of the human *COL2A1* gene (Barsh et al., 1984; Sangiorgi et al., 1985; Stoker et al., 1985; Elima et al., 1987; D'Alessio et al., 1988; Ala-Kokko and Prockop, 1990a,b; Chu and Prockop, 1993) linked to either a *COL1A1* promoter or a *COL2A1* promoter. The *COL2A1* gene was inserted into each construct as

‡ To whom correspondence should be addressed.

a 14 kb *SphI*–*SphI* fragment and a 12 kb *SphI*–*SphI* fragment isolated from a genomic cosmid library (Ala-Kokko and Prockop, 1990a,b). Both the constructs were assembled in the previously modified pJB8 cosmid vector (Ala-Kokko and Prockop, 1990a,b).

In the *COL1A1* promoter construct, the 5' end of the construct was a 2 kb *SphI*–*HindIII* fragment from the human *COL1A1* gene which extended from –500 to +1445 bp of the *COL1A1* gene (Barsh et al., 1984; D'Alessio et al., 1988; Chu and Prockop, 1992). The 3' end of the construct was a 3.5 kb *SphI*–*SphI* fragment which extended beyond the 3' end of the *COL2A1* gene (Ala-Kokko and Prockop, 1990a,b). For assembly of the construct, the *SphI* site at the 5' end of the 2 kb *SphI*–*HindIII* fragment from the *COL1A1* gene and the *SphI* site at the 3' end of the 3.5 kb *SphI*–*SphI* fragment from the *COL2A1* gene were converted into *SalI* sites. The *HindIII* site of the 2 kb *SalI*–*HindIII* fragment was then converted into a *SphI* site. In the first assembly step, the 2 kb *SalI*–*SphI* from the *COL1A1* gene and the 3.5 kb *SphI*–*SalI* fragment from the *COL2A1* gene were assembled by three-way ligation into *SalI* site of the modified pJB8 cosmid vector. In the second assembly step, the 14 and 12 kb *SphI*–*SphI* fragments from the *COL2A1* gene were inserted by three-way ligation into the *SphI* site of the construct obtained in the first step.

In the *COL2A1* promoter construct, the 5' end of the construct was an 8 kb *XbaI*–*SphI* fragment of the human *COL2A1* gene obtained from a KFE₁ cosmid clone (Sangiorgi et al., 1985). For the assembly of the construct, the *XbaI* site at the 5' end of the 8 kb *XbaI*–*SphI* fragment and the *SphI* site at the 3' end of the 12 kb *SphI*–*SphI* fragment were converted into *SalI* sites. The three fragments (8 kb *SalI*–*SphI*, 14 kb *SphI*–*SphI* and 12 kb *SphI*–*SalI*) were assembled by four-way ligation into the *SalI* site of the modified pJB8 cosmid vector (Ala-Kokko and Prockop, 1990a,b).

Cell culture and transfection

Human kidney tumour cells (HT1080; A.T.C.C., CCL 121) were grown in Dulbecco's modified Eagle's medium supplemented with 10% (v/v) fetal calf serum (Gibco/BRL) on 90 mm-diameter plastic tissue culture dishes. The procollagen gene constructs were cleaved from the vector with *SalI* without isolation of the gene from the vector. The neomycin-resistance gene construct (Law et al., 1983) was linearized with *BamHI*. For transfection, the calcium phosphate precipitation method (Sambrook et al., 1989) was used with a ratio of procollagen gene to neomycin-resistance gene of 200:1. Cells were incubated for 6 h with the DNA precipitate in 5 ml of medium containing from 8 to 12 µg per dish of the procollagen gene construct (11–16 µg of total DNA) and then subjected to a 3 min glycerol shock. Transfected cells were selected by growing them in the presence of G418 sulphate (Gibco/BRL) at a concentration of 1 mg/ml (potency over 450 µg/mg) for about 2 weeks.

Screening of transfected clones

G418-resistant clones were transferred to tissue culture plates with six 9.6 cm² wells. Fresh medium was added to confluent cells and after 24 h, proteins in the medium were precipitated with 5% (w/v) poly(ethylene glycol). Samples were centrifuged at about 15000 g for 30 min at 4 °C, and supernatants were discarded. Pellets were resuspended in 32 µl of storage buffer (0.4 M NaCl, 25 mM EDTA and 0.04% NaN₃ in 0.1 M Tris/HCl buffer at pH 7.4). Samples were treated with SDS and separated

by electrophoresis in 6% polyacrylamide gels in a mini-gel apparatus (Bio-Rad). Western-blot analysis was carried out with rabbit anti-(human procollagen II) antibodies (Ala-Kokko et al., 1991) and secondary antibodies of anti-rabbit IgG alkaline phosphatase (Promega). The antibodies were prepared (Ala-Kokko et al., 1991) with a 23-residue synthetic peptide that had the sequence: CAIDMSAFAGLGPREKGPDPLOY. Positive clones that secreted procollagen II were harvested and stored in liquid nitrogen.

Isolation and purification of procollagen II

Clones that secreted procollagen II were grown to confluence in 175 cm² cell-culture flasks under standard conditions without G418. To label the procollagen, the cell layer was washed with PBS and serum-free media containing 1 µCi of a uniformly ¹⁴C-labelled mixture of L-amino acids/ml (NEN/Du Pont), and 50 µg/ml ascorbic acid was added. After 24 h, the medium was removed and replaced with fresh medium of the same composition. After a second 24 h, the procedure was repeated under the same conditions except that the medium used for the third 24 h period did not contain ¹⁴C-labelled amino acids. The three 24 h samples of medium were pooled and mixed with 0.25 vol. of 5 × buffer to provide a final concentration of 25 mM EDTA and 0.04% NaN₃ in 0.1 M Tris/HCl buffer (pH 7.4). The sample was filtered through a 1.6 µm glass-fibre filter (Millipore), and proteins were precipitated overnight at 4 °C with 176 mg/ml (NH₄)₂SO₄. The precipitate was separated by centrifugation at 15000 g for 1 h at 4 °C. The pellets were resuspended overnight at 4 °C in the storage buffer, and insoluble material was removed by centrifugation. The supernatant was dialysed twice against 200 vol. of DEAE-cellulose column I buffer (2 M urea, 0.2 M NaCl, 5 mM EDTA and 0.04% NaN₃ in 0.1 M Tris/HCl buffer at pH 7.4). The sample was chromatographed on a DEAE-cellulose column (1.5 cm × 5 cm) equilibrated and eluted with the sample buffer (Peltonen et al., 1980; Fiedler-Nagy et al., 1982), and the flow-through fraction was collected.

The sample was dialysed against 200 vol. of DEAE-cellulose column II buffer (2 M urea, 2 mM EDTA and 0.04% NaN₃ in 0.075 M Tris/HCl buffer at pH 7.8), and chromatographed on the second DEAE-cellulose column as described previously (Peltonen et al., 1980; Fiedler-Nagy et al., 1982). Procollagen II and collagen IV were collected in the flow-through fraction and loaded on to a column (1.5 cm × 2 cm) of Q-Sepharose (Pharmacia) in the same buffer. Collagen IV was collected in the void volume. The procollagen II was eluted from the Q-Sepharose column with 0.4 M NaCl and immediately dialysed against 200 vol. of storage buffer before concentration to about 0.5 mg/ml on a membrane filter (Amicon YM-100) under a pressure of 48.3 kPa.

The proteins were analysed by electrophoresis in SDS using 6% polyacrylamide gels. Radiolabelled proteins were visualized with a phosphor storage imager (PhosphorImager; Molecular Dynamics) or by staining with Coomassie Brilliant Blue (Sigma). Specific radioactivity of the ¹⁴C-labelled procollagen II was determined by the hydroxyproline assay (Woessner, 1961) and by assuming 9.4% of hydroxyproline by weight for type-II procollagen (Fiedler-Nagy et al., 1982). The values ranged from about 1300 to 1500 c.p.m./µg.

cDNA sequencing, amino acid composition and carbohydrate content

Total RNA was isolated from the cells secreting procollagen II by guanidinium isothiocyanate/CsCl and density-gradient centri-

Figure 1 Schematic of the two gene constructs

(a) Construct containing the *COL2A1* gene with the *COL1A1* promoter. (b) Construct of the *COL2A1* gene with the *COL2A1* promoter.

fugation (Sambrook et al., 1989). Total RNA was used to synthesize cDNA (Gubler and Hoffman, 1983). For nucleotide sequencing, double-stranded DNAs from the phagemid were sequenced by the dideoxy method with T7 DNA polymerase (Sanger et al., 1977).

The amino acid composition was determined for us by the Wistar Protein Microchemistry Core Facility, Philadelphia, PA, U.S.A. The carbohydrate content of the pCcollagen II was determined by chromatographic methods (Anumula and Taylor, 1991) developed for the separation and characterization of acidic (sialylated) and neutral (asialo-complex and high-mannose) oligosaccharides released from glycoproteins with peptide *N*-glycosidase F, endo- β -*N*-acetylglucosaminidase F and endo- β -*N*-acetylglucosaminidase H using a carbohydrate analyser (Dionex BioLC). All the carbohydrate separations were carried out on a polymetric pellicular anion-exchange column HPIC-AS6/CarboPac PA-1 (Dionex) using only two eluents, namely, 0.5 M NaOH and 3% acetic acid/NaOH, pH 5.5, which were mixed with water to generate various gradients. Developed conditions for quantitative detection of carbohydrates with pulsed amperometry were necessary to obtain steady baselines at 0.1–0.3 μ A output with suitable sensitivity (< 5 pmol) in separations employing a variety of acidic and alkaline sodium acetate gradients.

Thermal-stability assay

14 C-labelled procollagen II was digested to collagen II by pepsin, and the thermal stability of the triple helix was assayed by rapid digestion with trypsin and chymotrypsin (Bruckner and Prockop, 1982). In brief, the samples were dissolved in 0.4 M NaCl plus 0.1 M Tris/HCl buffer at pH 7.4. They were preincubated at the indicated temperatures for 6 min and then digested for 2 min with 100 μ g/ml trypsin and 250 μ g/ml chymotrypsin (Sigma). The digestion was terminated by addition of 0.1 vol. of 5 mg/ml soya-bean trypsin inhibitor (Sigma). The samples were dissolved in hot sample buffer, heated at 100 °C for 2 min, and assayed by electrophoresis in SDS on a 6% polyacrylamide gel. 14 C-labelled collagen II isolated from chick sterna (Curran and Prockop, 1982, 1984) was used as a control.

Cleavage with procollagen N- and C-proteinases

Purified 14 C-labelled procollagen II was used as a substrate for procollagen N-proteinase and procollagen C-proteinase purified from chick embryo tendons (Hojima et al., 1985, 1989). The assay system was 40 μ l containing 10 μ g/ml procollagen, 150 mM

NaCl, 5 mM CaCl_2 and 2 units of enzyme in 50 mM Tris/HCl buffer at pH 7.4. The procollagen was digested with the N-proteinase at 32 °C for 5 h and with the C-proteinase at 37 °C for 5 h, and the digestion products were assayed by PAGE (Hojima et al., 1985, 1989).

Copy number of the type-II gene construct in the clones

To measure the copy number of the exogenous gene construct, genomic DNA was isolated from several independent clones of transfected cells synthesizing procollagen II. The DNA was isolated by chloroform/phenol extraction and ethanol precipitation on a workstation (Genepure 341; Applied Bio-Systems). The purified DNA was digested with *Eco*RI and used for Southern-blot analysis. For analysis of the construct containing the *COL1A1* promoter, the filter was probed with the 3.5 kb *Sph*I–*Sal*I fragment from the 3' end of the construct (Figure 1) which was labelled with [32 P]dCTP using a random priming kit (Stratagene). The probe hybridized with a 4.3 kb *Eco*RI band from the endogenous *COL2A1* gene and a 6.6 kb *Eco*RI band from the 3' end of the exogenous *COL2A1* gene. The filters were analysed on a phosphor storage imager, and the ratio of the 6.6 kb band to the 4.3 kb band was used to estimate the copy number of the transfected gene. For analysis of cells transfected with the construct containing the *COL2A1* promoter, the filter was probed with the 8 kb *Sal*I–*Sph*I fragment from the 5' end of the construct (Figure 1). The probe detected a 4.8 kb *Eco*RI fragment from the endogenous *COL2A1* gene and a single larger *Eco*RI fragment from each cell clone that included part of the gene construct and part of a flanking sequence of endogenous DNA.

RESULTS

Stably transfected cells

In the first series of experiments, a construct of the human *COL2A1* gene with the *COL1A1* promoter (Figure 1) was used to transfect HT1080 cells. The promoter fragment included about 0.5 kb of the 5' flanking region, the first exon and most of the first intron (Ala-Kokko et al., 1991). The promoter fragment was ligated through an *Sph*I site in intron 1B in the *COL2A1* gene to fragments that contained most of the human *COL2A1* gene and that extended about 3.5 kb beyond the second polyadenylation signal. About 100 clones resistant to G418 from each of three separate transfection experiments were analysed (Table 1). From 97 to 110 clones resistant to G418 from three separate transfections were grown in 9.6 cm² wells, and the

Table 1 Clones resistant to G418 and clones secreting procollagen II

Medium from clones grown in 9.6 cm² wells was assayed for secretion of procollagen II by Western blotting as shown in Figure 2.

Promoter in <i>COL2A1</i> gene construct	Experiment	Clones	
		G418- resistant	Synthesizing procollagen II
<i>COL1A1</i>	1	97	2
	2	107	1
	3	110	3
<i>COL2A1</i>	1	67	0
	2	73	2
	3	116	2

Figure 2 Analysis of proteins secreted from HT1080 cells after transfection with the *COL2A1* gene constructs

The assay was Western blot of proteins in the cell medium precipitated with $(\text{NH}_4)_2\text{SO}_4$. Lane 1, untransfected cells; lane 2, cells transfected with the construct containing the *COL1A1* promoter; lane 3, cells transfected with the construct containing the *COL2A1* promoter.

Figure 3 Analysis of ^{14}C -labelled proteins in the cell medium precipitated with $(\text{NH}_4)_2\text{SO}_4$

Proteins separated by SDS/PAGE were visualized with a phosphor storage imager. Lane 1, untransfected cells; lane 2, cells transfected with construct containing the *COL1A1* promoter; lane 3, cells transfected with construct containing the *COL2A1* promoter.

medium was assayed for secretion of procollagen II by Western blotting (Figure 2) and analysis of radiolabelled proteins (Figure 3). From one to three per 100 clones resistant to G418 were found to secrete procollagen II (Table 1).

In subsequent experiments, a similar construct containing the promoter of the *COL2A1* gene was employed. The construct contained about 4 kb of the promoter region and the complete human *COL2A1* gene ending 80 bp beyond the second polyadenylation site (Figure 1). From 67 to 116 clones resistant to G418 from each of three separate transfections were analysed (Table 1). Again, about two per 100 clones resistant to G418 were found to synthesize and secrete procollagen II.

Purification of the recombinant procollagen II

To purify the recombinant procollagen II, clones of cells synthesizing the protein were grown in about one hundred 175 cm² flasks. Proteins secreted into the medium by the cells were then purified chromatographically. As noted in Figure 3, the medium proteins included fibronectin and the $\alpha 1(\text{IV})$ and $\alpha 2(\text{IV})$ chains of collagen IV. As an initial step in the purification, the proteins were passed through a DEAE-cellulose column equilibrated with buffer containing 0.2 M NaCl (Peltonen et al., 1980; Fiedler-Nagy et al., 1982). The proteins in the medium were all eluted in the flow-through fraction and the column was employed primarily to remove unlabelled anionic materials that interfered with subsequent chromatographic separations (Fiedler-Nagy et

Figure 4 Purification of procollagen II from the culture medium of a clone of cells transfected with the construct containing the *COL1A1* promoter

Proteins were analysed by SDS/PAGE and staining with Coomassie Brilliant Blue R-250. Lane 1, medium proteins precipitated with 176 mg/ml $(\text{NH}_4)_2\text{SO}_4$; lane 2, proteins eluted from the second DEAE-cellulose column with 0.08 M NaCl; lane 3, proteins in flow-through fraction from the second DEAE-cellulose column; lane 4, proteins in void volume from Q-Sepharose column; lane 5, procollagen II eluted from Q-Sepharose column with 0.4 M NaCl.

al., 1982). In a second step, the proteins were chromatographed on a column of DEAE-cellulose equilibrated with 2 M urea and 2 mM EDTA in 75 mM Tris/HCl buffer, pH 7.8. Fibronectin and other proteins bound to the column but collagen IV and procollagen II were eluted in the flow-through fraction. In a third chromatographic step, the proteins recovered in the flow-through fraction from the second DEAE-cellulose column were chromatographed on a column of Q-Sepharose. The procollagen II was tightly bound by the column and was eluted with 0.4 M NaCl. The recombinant procollagen II was recovered as a single protein (lane 5 in Figure 4).

As assayed by hydroxyproline content, the yield of purified procollagen II was about 1 mg from about one hundred 175 cm² flasks. The recovery through the purification steps was about 50% and the initial concentration of procollagen II in the culture media was estimated to be about 0.6 mg/l.

Characterization of the protein by cDNA sequencing, amino acid composition and carbohydrate content

To characterize the recombinant protein, RNA was isolated from the clones used to prepare the recombinant procollagen II, the mRNA was transcribed into cDNA and the cDNA was sequenced. Sequencing of all the bases encoding the pro $\alpha 1(\text{II})$ chain demonstrated that the nucleotide sequence was the same as described previously except for a single base substitution in the N-propeptide which converted the codon (-CCT-) for proline at positions +72 in the prepro $\alpha 1(\text{II})$ chain into -CAT-, the codon for histidine.

To characterize the recombinant protein further, procollagen II secreted by the cells was purified, and converted into collagen II by cleavage with procollagen N- and C-proteinases. Collagen II was precipitated by centrifugation, and its amino acid composition was determined. The observed amino acid composition was consistent with the values for amino acid composition of collagen II reported for pepsin-treated samples of the protein (Piez, 1984, and Table 2). It was also consistent with the predicted amino acid composition from the cDNA sequence.

In parallel experiments, recombinant procollagen II was cleaved to pCcollagen II with procollagen N-proteinase. The resulting pCcollagen II was concentrated by ultrafiltration and repurified on a Q-Sepharose column. The carbohydrate content was determined using a carbohydrate analyser. The results (Table 3) demonstrated that the galactose and glucose contents were less

Table 2 Amino acid composition of recombinant collagen II

The results for recombinant collagen II are means of three analyses. N.D., not determined.

	Composition (residues/1000)		
	Pepsin-treated collagen II*	Encoded by cDNA	Recombinant collagen II ^b
4-Hyp	97	113†	91
Asp	43	43	49
Thr	23	21	29
Ser	25	28	34
Glu	89	92	91
Pro	120	111	102
Gly	333	331	290
Ala	103	100	99
Half-Cys	0	0	< 3
Val	18	18	27
Met	10	10	13
Ile	9	11	18
Leu	26	26	42
Tyr	2	2	9
Phe	13	13	24
Hyl	20	25†	N.D.
Lys	15		27
His	2	2	8
Arg	50	51	54

* Values from Piez (1984).

† Assuming Y-position Pro and Lys are fully hydroxylated.

Table 3 Carbohydrate content of recombinant pCcollagen II

Sugar	Content (residues per chain)	
	Collagen II*	Recombinant pCcollagen II
Galactose	16	5.1
Glucose	12	9.4
Mannose	0	2.5†
N-Acetylglucosamine	0	3†
Fucose	0	2.1†

* Values from Piez (1984).

† Residues expected to be present in the asparagine-linked oligosaccharide unit composed of N-acetylglucosamine and mannose found in the C-propeptide (see the text).

than the values reported for $\alpha 1(\text{II})$ chains isolated from cartilage (Piez, 1984). As the glucose in collagen II and procollagen II is found as glucosyl- $\alpha 1,2$ -galactosyl- β -hydroxylysine, the results demonstrated that some of the hydroxylysyl residues were glycosylated but they were not overmodified by the post-translational reactions. The presence of mannose, N-acetylglucosamine and fucose is consistent with previous indications that the C-propeptide of fibrillar procollagens contains a mannose-rich oligosaccharide (see Clark and Kefalides, 1976).

Assays of conformational integrity of the recombinant procollagen II

The conformational integrity of the recombinant procollagen II was assayed by a series of protease digestions.

Figure 5 Assay of thermal stability of triple helix by brief proteinase digestion of ¹⁴C-labelled recombinant collagen II

(a) Standard collagen II from chick embryos. (b) Collagen II from cells transfected with construct containing the *COL1A1* promoter. (c) Collagen II from cells transfected with the construct containing the *COL2A1* promoter.

Figure 6 Cleavage of ¹⁴C-labelled recombinant procollagen II with procollagen N-proteinase and C-proteinase

Lane 1, undigested procollagen II; lane 2, procollagen II cleaved with C-proteinase; lane 3, procollagen II cleaved with N-proteinase; lane 4, recombinant procollagen II incubated without proteinase at 37 °C for 5 h.

In one series of experiments, the thermal stability of the triple helix was assayed by brief protease digestion with high concentrations of trypsin and chymotrypsin (Bruckner and Prockop, 1982). As indicated in Figure 5, the triple helix of recombinant collagen II had an apparent melting temperature of about 41 °C, a value that was the same as that of control collagen II isolated from the sterna of chick embryos, and consistent with previously reported values for collagen II (Bächinger and Morris, 1990). Also, no difference in apparent melting temperature was seen between procollagen II purified from a clone of cells transfected with the two promoter constructs.

In further experiments, recombinant procollagen II was tested as substrate for procollagen N-proteinase and procollagen C-proteinase from chick embryo tendons (Figure 6). The sizes of the fragments generated by digestion with the N-proteinase were consistent with the conclusion that cleavage occurred at the single N-proteinase-cleavage site in the protein so that pC α chains of pCcollagen were generated (lane 3). The sizes of the fragments generated by digestion with the C-proteinase were consistent with the conclusion that cleavage occurred at the

Table 4 Copy number and secretion of procollagen II by clones of transfected cells

Values for the ratio of procollagen II to collagen IV in the medium were obtained by ^{14}C -labelling of clones grown on 9.6 cm² plates, $(\text{NH}_4)_2\text{SO}_4$ precipitation of the medium proteins, separation of the proteins by SDS/PAGE and assay of the ^{14}C -labelled bands on a phosphor storage imager (see the text and Figure 3).

Promoter	Experiment	Copy number	Ratio of procollagen II to collagen IV in medium
<i>COL1A1</i>	1	8	6.5
	2	2	1.8†
	3	1	0.6
<i>COL2A1</i>	1	2	2.5*
	2	1	1.1
	3	1	0.9

* Clone used for purification of recombinant procollagen II (Figure 4).

† Clones used to assay thermal stability of triple helix (Figure 5).

single C-proteinase-cleavage site in the protein so that pN α chains of pNcollagen were generated (lane 2).

Gene copy number of expressing clones

To assay the copy number of the exogenous gene in the transfected cells, Southern-blot assays were performed under conditions in which the same probe detected distinct *Eco*RI fragments from the endogenous *COL2A1* gene and the exogenous gene constructs. As indicated in Table 4, the copy number varied from 1 to about 8 in different clones of transfected cells. With both gene constructs, the copy number was closely related to the level of expression as estimated by the relative amount of recombinant procollagen II and endogenous collagen IV secreted into the media.

DISCUSSION

Expression of the gene for human procollagen II in a recombinant system is important for several reasons. One reason is that purification of either normal procollagen II or normal collagen II from tissues has proven to be difficult. Small amounts of collagen II can be extracted from cartilage, but the yields of protein are small (see Lee and Piez, 1983; Curran and Prockop, 1984; Bächinger and Morris, 1990). Also, the extracted collagen II is frequently partially cleaved in the telopeptide regions, and monomer preparations isolated from the extracts are frequently contaminated by covalently cross-linked oligomers (Lee and Piez, 1983). Partial cleavage of the telopeptides and the presence of oligomers has made it difficult to obtain reproducible data on the self-assembly of collagen II into fibrils. Isolation of homogeneous preparations of procollagen II have proved to be equally difficult. Small amounts of the protein can be obtained from cell cultures of chondrocytes (Curran and Prockop, 1984) but chondrocytes are poorly amplified in culture, and they frequently dedifferentiate so as to cease synthesis of procollagen II and other cartilage-specific macromolecules (Elima and Vuorio, 1989; Aulthouse et al., 1989). Therefore recombinant systems to generate significant amounts of procollagen II are important for defining the normal structure-function relationships of the protein. A further reason for interest in recombinant procollagen

II is that mutations in the gene for procollagen II have been implicated as a cause of several human diseases, including a series of chondrodysplasias and a subset of primary generalized osteoarthritis (Francomano et al., 1987; Palotie et al., 1989; Ala-Kokko et al., 1990; Anderson et al., 1990; Tiller et al., 1990; Vissing et al., 1990; Ahmad et al., 1991; Kuivaniemi et al., 1991). Establishing the causal relationship between mutations in the gene and the biological function of the protein, however, has proved difficult because of the inaccessibility of the human protein.

The system for expressing the human *COL2A1* gene described here overcomes most of these problems. Expression of the *COL2A1* gene in HT1080 cells made it possible to purify the protein from culture medium, and milligram quantities of the protein were obtained. The protein was shown to have a native conformation in that the triple helix had a normal thermal stability as assayed by brief protease digestion. Also, it was appropriately cleaved by procollagen C-proteinase and, most importantly, by procollagen N-proteinase which has a specific requirement for a procollagen substrate with a native conformation (see Hojima et al., 1989). Therefore the system can provide adequate amounts of native recombinant procollagen II for studying self-assembly of collagen II into fibrils by cleavage of the procollagen with specific proteinases (Kadler et al., 1990). Also, the potential for expressing mutated *COL2A1* genes in the same system makes it possible to examine the effects of mutations that change the primary structure of the protein on its biological function (Torre-Blanco et al., 1992).

In previous experiments on expression of the *COL2A1* gene in mouse NIH 3T3 cells (Ala-Kokko et al., 1991), the constructs contained about 0.5 kb of the promoter region, the first exon and most of the first intron of the *COL1A1* gene. As NIH 3T3 cells normally synthesize procollagen I, it was assumed (Ala-Kokko et al., 1991) that expression of the procollagen II gene was obtained because the construct contained a promoter from a gene normally expressed in the same cells. The results here, however, demonstrate that a promoter of an endogenously expressed gene is not essential to obtain expression in stably transfected cells. Instead, the site of integration is probably more critical than the structure of the promoter. HT1080 cells synthesize collagen IV and therefore they contain the post-translational enzymes required for hydroxylation and glycosylation of collagens (Tryggvason et al., 1980; Prockop and Kivirikko, 1984; Ala-Kokko et al., 1991). This feature of the cells is probably essential to obtain expression of a collagen or procollagen gene. However, clones synthesizing procollagen II were obtained with constructs containing either a *COL1A1* promoter or a *COL2A1* promoter, even though the endogenous genes for these fibrillar procollagens are not expressed in HT1080 cells. Moreover, only 0.5 kb of the promoter was present in the *COL1A1* promoter construct even though previous experiments with transient cell transfection and transgenic mice suggested 2.5 kb or more of the *COL1A1* promoter may be necessary for correct expression of the gene (see Boast et al., 1990; Slack et al., 1991; Pavlin et al., 1992). In addition, the results with both constructs indicated that the level of expression in clones found to synthesize the recombinant protein was closely related to gene copy number. In previous experiments with the *COL1A1* promoter construct (Olsen et al., 1991; Ala-Kokko et al., 1991), no relationship was seen between copy number and level of expression. The difference from the observations here was that randomly selected G418-resistant clones were compared in the previous experiments, whereas here only the subset of G418-resistant clones that secreted procollagen II were compared. Overall therefore the results indicate that if stably transfected cells are screened for clones expressing an

exogenous gene, a subset of clones can be obtained that express the gene in a copy-dependent manner even if it contains a constitutively inactive promoter. Within the content of such experiments, the site of stable integration of the foreign gene is apparently far more important than the *cis*-regulatory elements.

We are indebted to Dr. Yoshio Hojima for the supply of procollagen N-proteinase, procollagen C-proteinase and procollagen II from the sterna of chick embryos. This work was supported in part by grants from the National Institutes of Health (AR38188), the March of Dimes-Birth Defects Foundation, and the Lucille P. Markey Charitable Trust.

REFERENCES

- Ahmad, N. N., Ala-Kokko, L., Knowlton, R. G., Weaver, E. J., Maguire, J. I., Tasman, W. and Prockop, D. J. (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88**, 6624–6627
- Ala-Kokko, L. and Prockop, D. J. (1990a) *Matrix* **10**, 279–284
- Ala-Kokko, L. and Prockop, D. J. (1990b) *Genomics* **8**, 454–460
- Ala-Kokko, L., Baldwin, C. T., Moskowitz, R. W. and Prockop, D. J. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 6565–6568
- Ala-Kokko, L., Hyland, J., Smith, C., Kivirikko, K. I., Jimenez, S. A. and Prockop, D. J. (1991) *J. Biol. Chem.* **266**, 14175–14178
- Anderson, I. J., Goldberg, R. B., Marion, R. W., Upholt, W. B. and Tsipouras, P. (1990) *Am. J. Hum. Genet.* **46**, 896–901
- Anumula, K. R. and Taylor, P. B. (1991) *Eur. J. Biochem.* **195**, 269–280
- Aulthouse, A. L., Beck, M., Griffey, E., Sanford, J., Arden, K., Machado, M. A. and Horton, W. A. (1989) *In Vitro Dev. Biol.* **25**, 659–668
- Bächinger, H. P. and Morris, N. P. (1990) *Matrix* **10**, 331–338
- Barsh, G. S., Roush, C. L. and Gelinas, R. E. (1984) *J. Biol. Chem.* **259**, 14906–14913
- Boast, S., Su, M.-W., Ramirez, F., Sanchez, M. and Avvedimento, E. V. (1990) *J. Biol. Chem.* **265**, 13351–13356
- Bruckner, P. and Prockop, D. J. (1982) *Anal. Biochem.* **110**, 360–368
- Chu, M.-L. and Prockop, D. J. (1993) in *Extracellular Matrix and Inheritable Disorders of Connective Tissue* (Royce, P. M. and Steinmann, B., eds.) pp. 149–166, Alan R. Liss, New York
- Clark, C. C. and Kefalides, N. A. (1976) *Proc. Natl. Acad. Sci. U.S.A.* **73**, 34–38
- Curran, S. and Prockop, D. J. (1982) *Biochemistry* **21**, 1482–1487
- Curran, S. and Prockop, D. J. (1984) *Biochemistry* **23**, 741–746
- D'Alessio, M., Bernard, M., Pretorius, P. J., de Wet, W. and Ramirez, F. (1988) *Gene* **67**, 105–115
- Elima, K. and Vuorio, E. (1989) *FEBS Lett.* **258**, 195–198
- Elima, K., Vuorio, T. and Vuorio, E. (1987) *Nucleic Acids Res.* **15**, 9499–9504
- Fiedler-Nagy, C., Bruckner, P., Hayashi, T., Fietzek, P. and Prockop, D. (1982) *J. Biol. Chem.* **257**, 9181–9188
- Francomano, C. A., Liberfarb, R. M., Hirose, T., Maumenee, I. K., Streeten, H. A., Meyers, D. A. and Pyritz, R. E. (1987) *Genomics* **1**, 293–296
- Gubler, U. and Hoffman, B. J. (1983) *Gene* **25**, 263–269
- Hojima, Y., van der Rest, M. and Prockop, D. J. (1985) *J. Biol. Chem.* **260**, 15996–16003
- Hojima, Y., McKenzie, J., van der Rest, M. and Prockop, D. J. (1989) *J. Biol. Chem.* **264**, 11336–11345
- Kadler, K. E., Hojima, Y. and Prockop, D. J. (1990) *Biochem. J.* **268**, 339–343
- Kuivaniemi, H., Tromp, G. and Prockop, D. J. (1991) *J. Clin. Invest.* **88**, 1441–1444
- Law, M.-F., Byrne, J. C. and Howley, P. M. (1983) *Mol. Cell. Biol.* **3**, 2110–2115
- Lee, S. L. and Piez, K. A. (1983) *Collagen Relat. Res.* **3**, 89–103
- Lee, S. T., Smith, B. D. and Greenspan, D. S. (1989) *J. Biol. Chem.* **264**, 20683–20687
- Olsen, A. S., Geddis, A. E. and Prockop, D. J. (1991) *J. Biol. Chem.* **266**, 1117–1121
- Palotie, L., Väimäisen, P., Ott, J., Ryhänen, L., Elima, K., Vikkula, M., Cheah, K., Vuorio, E. and Peltonen, L. (1989) *Lancet* **i**, 924–927
- Pavlin, D., Lichtler, A. C., Bedalov, A., Kream, B. E., Harrison, J. R., Thomas, H. F., Gronowicz, G. A., Clark, S. H., Woody, C. O. and Rowe, D. W. (1992) *J. Cell Biol.* **116**, 227–236
- Piez, K. A. (1984) in *Extracellular Matrix Biochemistry* (Piez, K. A. and Reddi, A. H., eds.), pp. 49–50, Elsevier, New York
- Peltonen, L., Palotie, A., Hayashi, T. and Prockop, D. J. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 162–166
- Pihlajaniemi, T., Myllylä, R., Alitalo, K., Vaheeri, A. and Kivirikko, K. I. (1981) *Biochemistry* **20**, 7409–7415
- Prockop, D. J. and Kivirikko, K. I. (1984) *N Engl. J. Med.* **311**, 376–386
- Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edn., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
- Sanger, F., Nicklen, S. and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467
- Sangiorgi, F. O., Benson-Chanda, V., de Wet, W. J., Sobel, M. E., Tsipouras, P. and Ramirez, F. (1985) *Nucleic Acids Res.* **13**, 2207–2205
- Schnieke, A., Dziadek, M., Bateman, J., Mascara, T., Habers, K., Gelinas, R. and Jaenisch, R. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 8869–8873
- Slack, J. L., Liska, D. J. and Bornstein, P. (1991) *Mol. Cell. Biol.* **11**, 2066–2074
- Stoker, N. G., Cheah, K. S. E., Griffin, J. R., Pope, F. M. and Solomon, E. (1985) *Nucleic Acids Res.* **13**, 4613–4622
- Tiller, G. E., Rimoim, D. L., Murray, L. W. and Cohn, D. H. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 3889–3893
- Torre-Blanco, A., Adachi, E., Romanic, A. M. and Prockop, D. J. (1992) *J. Biol. Chem.* **267**, 4968–4973
- Tryggvason, K., Gehron Robey, P. and Martin, G. R. (1980) *Biochemistry* **19**, 1284–1289
- Vissing, H., D'Alessio, M., Lee, B., Ramirez, F., Godfrey, M. and Hollister, D. W. (1990) *J. Biol. Chem.* **264**, 18265–18267
- Woessner, J. (1961) *Arch. Biochem. Biophys.* **93**, 440–447