

# Expression of Type XXIII Collagen mRNA and Protein<sup>\*[5]</sup>

Received for publication, May 1, 2006 Published, JBC Papers in Press, May 25, 2006, DOI 10.1074/jbc.M604131200

Manuel Koch<sup>†§¶</sup>, Guido Veit<sup>‡</sup>, Sigmar Stricker<sup>||</sup>, Pinaki Bhatt<sup>\*\*</sup>, Stefanie Kutsch<sup>‡</sup>, Peihong Zhou<sup>\*\*</sup>, Elina Reinders<sup>‡</sup>, Rita A. Hahn<sup>\*\*</sup>, Rich Song<sup>\*\*</sup>, Robert E. Burgeson<sup>1</sup>, Donald R. Gerecke<sup>\*\*</sup>, Stefan Mundlos<sup>||¶‡</sup>, and Marion K. Gordon<sup>\*\*2</sup>

From the <sup>‡</sup>Center for Biochemistry, <sup>§</sup>Department of Dermatology, and <sup>¶</sup>Center for Molecular Medicine Cologne, University of Cologne, Joseph-Stelzmann Strasse 52, 50931 Cologne, Germany, <sup>||</sup>Development and Disease Group, Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany, <sup>¶¶</sup>Institute for Medical Genetics, University Medicine Charité, 13353 Berlin, Germany, and <sup>\*\*</sup>Department of Pharmacology and Toxicology, Ernest Mario School of Pharmacy and the Environmental and Occupational Health Sciences Institute, Rutgers University, Piscataway, New Jersey 08854

Collagen XXIII is a member of the transmembranous subfamily of collagens containing a cytoplasmic domain, a membrane-spanning hydrophobic domain, and three extracellular triple helical collagenous domains interspersed with non-collagenous domains. We cloned mouse, chicken, and human  $\alpha 1$ (XXIII) collagen cDNAs and showed that this non-abundant collagen has a limited tissue distribution in non-tumor tissues. Lung, cornea, brain, skin, tendon, and kidney are the major sites of expression. In contrast, five transformed cell lines were tested for collagen XXIII expression, and all expressed the mRNA. *In vivo* the  $\alpha 1$ (XXIII) mRNA is found in mature and developing organs, the latter demonstrated using stages of embryonic chick cornea and mouse embryos. Polyclonal antibodies were generated in guinea pig and rabbit and showed that collagen XXIII has a transmembranous form and a shed form. Comparison of collagen XXIII with its closest relatives in the transmembranous subfamily of collagens, types XIII and XXV, which have the same number of triple helical and non-collagenous regions, showed that there is a discontinuity in the alignment of domains but that striking similarities remain despite this.

Tissues use specific sets of collagens, often synthesized simultaneously, to achieve and maintain particular functional properties. Most collagens are secreted and assembled within the extracellular environment; however, a growing subclass of collagens are transmembranous and inserted into the plasma membrane in a type II orientation to extend their extracellular collagenous domains from the cell surface.

The subclass of transmembranous collagens currently

includes types XIII, XVII, XXIII, and XXV, summarized in a recent review (1). The group has also been referred to as the MACITs for membrane-associated collagens with interrupted triple helices (2). Type XIII collagen, the first member of the group identified (3), has an important function in muscle tissue. Engineered genetic mutations in the *Col13a1* gene in one case causes cardiovascular defects (4) and in another causes abnormal skeletal muscle myofibers with progressive myopathy that is worsened by exercise (5). Type XIII collagen mediates cell attachment through integrin  $\alpha 1\beta 1$  but does not interact through another common collagen receptor,  $\alpha 2\beta 1$  (6). The second transmembranous collagen to be identified was type XVII (7), which was known for many years as bullous pemphigoid antigen 2 and BP180 prior to the elucidation of its collagenous nature. Being a component of the hemidesmosome (8), type XVII collagen provides structural integrity to the cornea (9–11) and skin (12), clearly demonstrated by mutations in the *COL17A1* gene that cause epidermolysis bullosa (13, 14). Our laboratories added collagen XXIII to the family of transmembranous collagens by identifying a human EST<sup>3</sup> and using its sequence to clone a fragment of chicken collagen XXIII cDNA to examine its expression in cornea (15). We also cloned the full-length mouse collagen XXIII cDNA sequence and deposited it into GenBank<sup>TM</sup> (accession number AF410792). Cloning and expression of the rat and human  $\alpha 1$ (XXIII) cDNAs revealed evidence that (i) the molecule is shed from the cell surface by selective proteolysis suggesting an involvement of furin and (ii) that the expression is up-regulated with metastatic potential in prostate cancer cell lines (16). For collagen XXV, the newest member of the group, mRNA could be specifically detected in neurons and has been identified as a component of the senile plaques characteristic of Alzheimer disease (17).

Here we present the strategies for obtaining mouse, chicken, and human  $\alpha 1$ (XXIII) collagen cDNAs. The sequences and domain structures of the orthologs were compared as were the distribution and abundance of the mRNA in mature and developing tissues. The protein was examined in tissues using polyclonal antibodies generated in guinea pig and rabbit against the ectodomain of mouse collagen XXIII. The antibodies revealed

<sup>\*</sup> This work was supported by Deutsche Forschungsgemeinschaft Grant SFB 589 P1 (to M. K.), an award from the Alexander von Humboldt Foundation (to M. K.), the Federal Ministry of Education and Research in Germany (to M. K.), and Grant HL67708 (to D. R. G.), NIEHS Center Grant ES05022, and NEI Grant EY09056 (to MKG) from the National Institutes of Health. 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.

[5] The on-line version of this article (available at <http://www.jbc.org>) contains supplemental material.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank<sup>TM</sup>/EBI Data Bank with accession number(s) AY898961, AF410792, and AY876377.

<sup>1</sup> Emeritus of the Cutaneous Biology Research Center, Massachusetts General Hospital East, Harvard Medical School, Charlestown, MA 02129.

<sup>2</sup> To whom correspondence should be addressed: Dept. of Pharmacology and Toxicology, Ernest Mario School of Pharmacy, Rutgers University, 170 Frelinghuysen Rd., Piscataway, NJ 08854. Tel.: 732-445-3751; Fax: 732-445-0119; E-mail: magordon@eohsi.rutgers.edu.

<sup>3</sup> The abbreviations used are: EST, expressed sequence tag; Col, collagenous; NC, non-collagenous; dbEST, database of ESTs; FACIT, fibril-associated collagen with interrupted triple helices; HEK, human embryonic kidney; EBNA, Epstein-Barr nuclear antigen 1; RACE, rapid amplification of cDNA ends; RT, reverse transcription; TBS, Tris-buffered saline; E, embryonic day.

that some tissues contain mostly the full-length transmembranous form, whereas in other tissues the shed ectodomain is the major form of collagen XXIII.

## MATERIALS AND METHODS

**Identification and Cloning of  $\alpha 1$ (XXIII) Collagen cDNAs**—Stretches of amino acid sequences, derived from specific domains in the FACIT collagen types, were used as queries to perform BLAST searches (18) to pick up related, but unique, collagen cDNAs in the human dbEST (19). A unique clone (GenBank™ accession number W22262) was identified that encodes a portion of a triple helical domain and 18 amino acid residues of a non-collagenous domain with strong resemblance to type XIII collagen. The clone was obtained, completely sequenced, and assigned the next available number in the collagen family,  $\alpha 1$ (XXIII) (see supplemental material). In total, the composite full-length human  $\alpha 1$ (XXIII) collagen cDNA is 3,060 nucleotides and was assigned GenBank™ accession number AY898961.

The human sequence was used to search mouse ESTs, and clone AW323570 was identified. After extending the cDNA sequence by 5'-RACE using mouse lung cDNA, an overlap was found with a 5'-end EST (GenBank™ accession number AW494383). The primary structure of the entire mouse  $\alpha 1$ (XXIII) collagen mRNA was verified by PCR and sequence analysis and was submitted to GenBank™ under accession number AF410792. To obtain the initial chick cDNA, primers derived from nearly identical regions in the human and mouse sequences were used to amplify 13-day chick embryo corneal cDNA. A small chick  $\alpha 1$ (XXIII) collagen cDNA of ~100 bp was obtained and reported in the 2000 ARVO meeting abstracts issue of the journal *Investigative Ophthalmology and Visual Science* (15). The complete chick  $\alpha 1$ (XXIII) collagen cDNA was deposited in GenBank™ under accession number AY876377 (for cloning see supplemental material).

The ExPASy Sim program from the Search Launcher at the Baylor College of Medicine Human Genome Sequencing Center website (found at searchlauncher.bcm.tmc.edu/) was used to align polypeptide chains. For the comparison of the longest version of the human  $\alpha 1$ (XXIII) collagen polypeptide (accession number AY158895), human type XIII collagen (accession number AJ293624), and human type XXV collagen (accession number NM\_198721), placement of amino acid residues was biased to align conserved exons in each collagen type in an effort to reflect areas of conserved gene structure (see supplemental material). To accomplish this, the exon/intron structure of each gene was determined using GenBank™ tools.

**Tissues**—Human placenta, lung, and kidney total RNA were purchased from Clontech. Human cornea was from Lions Eye Bank of Oregon (Portland, OR), and human amnion was from Marie France Champlaud-Steiner (formerly of Cutaneous Biology Research Center, Massachusetts General Hospital). All tissues were immediately frozen in liquid nitrogen until RNA isolation, which was accomplished using TRIzol™ (Invitrogen) following the manufacturer's instructions. On-column DNase digestion was performed on the RNA as outlined in appendix D instructions in the RNeasy™ minikit handbook from Qiagen. Human donor corneal epithelial, stromal, and

endothelial cDNAs were kindly provided by Mitch Watsky, Sally Twining, and Kim Vaughn. Human immortalized corneal epithelial cells (20) (a gift from Dr. Fu-Shin Yu with permission of Dr. K. Araki-Sasaki), lung and esophageal cell lines A549 and HET-1A (a gift from Drs. C. S. Yang and L. Chen), esophageal cell line SEG-1 (from Drs. C. S. Yang and L. Chen with permission of Dr. David Beer), and the SKGT4 cell line (also from Drs. C. S. Yang, Luke Chen, and Xiaochun Xu with permission of Dr. David Schrupp) were grown in culture, then scraped from plates, and used for RNA isolation. Mouse cornea, lung, skin, and sterna were dissected for RNA isolation; bone and tendon were purchased from Pel-Freeze Biologicals. RNA from human cell lines and all mouse tissues was isolated using Qiagen RNeasy kits. For chick tissue, corneas were dissected from 7-, 9-, 11-, 13-, and 15-day embryos. With the exception of the 7-day corneas, epithelia were removed from the stromal/endothelial layer with EDTA and dispase, and RNA was isolated as described previously (21). For all tissues collected, RNA was quantified by absorbance measured at 260 nm.

**PCR, RT-PCR, and Relative RT-PCR**—Routine 30-cycle amplifications were as described previously (21). RACE reactions were as performed previously (22) with modifications as described by Koch *et al.* (23). With the exception of the human corneal epithelial, stromal, and endothelial mRNA, all other human relative RT-PCRs were performed in the linear range using the non-radioactive method described by Koch *et al.* (23). The human primers were AACATTCAAGAGTATGCC-CACC and CACACAGTGATCTCCATCTACACGG (944-bp product). Products were run on ethidium-containing agarose gels and photographed with an Eagle Eye II digital camera. Relative RT-PCR to quantitate mRNA abundance in mouse and chick tissues as well as in human corneal epithelium, stroma, and endothelium was done incorporating radioactive nucleotide into the product as described previously (23, 24). Human collagen XXIII primers were TAACATTCAAGAGTATGCC-CACC and AGGTTCTTACAGGGCAGTA (240-bp product). Mouse collagen XXIII primers were GACCTGGAGAACCAG-GACTCGATG and CATCCAAACGGATCTGTACAGGTC (247-bp product). Chick collagen XXIII primers were GGAG-AAATGGGCTTATCGGGT and CGGGTAAGCCAATGAG-CCCT (215-bp product). PCR was performed for 29 cycles in the linear amplification range for fragments. Bands were cut from gels, and radioactivity was counted. The collagen XXIII signal was normalized by dividing it by radioactivity counts of the attenuated 18 S product (Ambion primer to Competimer ratio, 1:20) or, for chick, by the radioactivity counts of the glyceraldehyde-3-phosphate dehydrogenase product. The resulting normalizations are called "adjusted cpm." Because all primers were tested for optimal amplification capability and samples for each species were amplified using one master mixture and run on the same gel, the histograms show the relative differences between the amplified products.

**Whole Mount and Tissue Section *In Situ* Hybridization**—Two probes were generated by RT-PCR (Clontech) using the following primer pairs: forward CCCAAGTTCAGCACG-CTTC and reverse T7-CCATGAATGGAGCAGGAGCTAG and forward TCCAAGGTCCCAAGGGCCTG and reverse T7-CTGTCCAGCCTCACCTGAGCC. The whole mount *in situ*

## $\alpha 1$ (XXIII) Collagen

hybridizations was done as described previously (25), and section *in situ* hybridizations were performed with digoxigenin-labeled riboprobes using a semiautomated TECAN Genesis robot platform applying the Genepaint system (26). Both probes stained equivalent structures.

**Recombinant Expression of  $\alpha 1$ (XXIII) Collagen in Human Embryonic Kidney (HEK) 293-EBNA Cells**—The full-length  $\alpha 1$ (XXIII) collagen cDNA could not be amplified in one piece so an AvrII site was incorporated without altering the amino acid sequence at positions 45–49. The 5'-fragment was amplified using a KpnI sequence linked to CTGAGTCTGAGCGGGCGGCTAGG as the sense primer and an AvrII sequence linked to GGTCCCGGGTAGCC as the antisense primer. The 3'-fragment was amplified with forward primer GGGACC linked to TCTGGACGGC through an AvrII sequence and, as reverse primer, an XhoI sequence linked to CTTATGCCAGCAACCAGGCACAGG. PCRs were done on mouse lung cDNA in 6% Me<sub>2</sub>SO using Long Expand DNA polymerase (Roche Diagnostics). The products were ligated (Rapid DNA ligation kit, Roche Diagnostics) into a modified pCEP-Pu vector carrying a 3' FLAG and His<sub>6</sub> tag. The ectodomain expression plasmid was generated by inserting the SPARC (secreted protein, acidic and rich in cysteine) signal peptide coding sequence (27) between the HindIII and NheI sites in pCEP-4 (provided by Ernst Poeschl) and then inserting the collagen XXIII ectodomain amplified from mouse lung cDNA. The sense primer was an NheI sequence linked to ACAGGCGGCCGCCCTGCATGGCC, and the antisense primer was an XhoI sequence linked to CTTATGCCAGCAACCAGGCACAGG. Cloned and expanded products were sequenced to verify their identities. Transfection experiments and protein purification were done as described previously (22). The vectors containing the collagen XXIII cDNAs were introduced into HEK 293-EBNA cells (Invitrogen) by transfection with the reagent FuGENE 6 (Roche Applied Science) according to the manufacturer's recommendations. The cells were selected with puromycin (1.25  $\mu$ g/ml), stably transfected HEK 293-EBNA cells were pseudo-subcloned, and the highest protein-producing clones were expanded.

**Use of Collagen XXIII Ectodomain for Antibody Production**—Antibodies against mouse collagen XXIII were produced as described previously (28). Briefly for large scale protein production, the ectodomain-transfected cells were transferred to serum free Dulbecco's modified Eagle's medium/F-12 with Glutamax<sup>TM</sup> (Invitrogen) supplemented with 250  $\mu$ M L-ascorbic acid and 450  $\mu$ M L-ascorbic acid 2-phosphate. The collected cell culture supernatant was supplemented with 1 mM phenylmethylsulfonyl fluoride (Sigma) and, after filtration, was passed over a gelatin-Sepharose column (GE Healthcare) before being applied to a TALON metal affinity resin column (BD Biosciences). The recombinant protein was eluted stepwise with Tris-buffered saline (TBS; 20 mM Tris-HCl, pH 8.0, 150 mM NaCl) containing 40–250 mM imidazole. After dialysis against TBS the protein was used for immunization of two rabbits and two guinea pigs following standard procedures. The antisera collected from animals weeks later were purified by affinity chromatography on a column with antigen coupled to CNBr-activated Sepharose (GE Healthcare). The specific antibodies

were eluted with 150 mM NaCl, 0.1 M triethylamine, pH 11.5, and the eluate was neutralized with 1 M Tris-HCl, pH 6.8.

**Solid Phase Assay**—The specificity of the antibodies against collagen XXIII was confirmed in solid phase assays. The ectodomains of collagens XIII and XXV and the unrelated His-tagged protein unc5h2 were expressed and purified using the same method as used for collagen XXIII ectodomain production described above. The purified proteins were diluted in TBS, pH 7.4, and 10  $\mu$ g/ml (500 ng/well) were coated overnight at 4 °C onto 96-well plates (Nunc Maxisorb). After washing with TBS, nonspecific binding sites were blocked with 5% skim milk powder in TBS for 2 h at room temperature. Serial dilutions of the affinity-purified anti-collagen XXIII antibodies in blocking buffer (1:300–1:100,000) were added to wells and incubated for 1.5 h. Bound primary antibodies were detected with secondary antibody that was either swine anti-rabbit horseradish peroxidase-coupled IgG (DakoCytomation) or rabbit anti-guinea pig horseradish peroxidase-coupled IgG (Sigma). For enzymatic reaction, 50  $\mu$ l/well 0.25 mM tetramethylbenzidine and 0.005% (v/v) H<sub>2</sub>O<sub>2</sub> in 0.1 M sodium acetate, pH 6.0 were incubated for 10 min. The reaction was stopped with 50  $\mu$ l/well 2.5 M H<sub>2</sub>SO<sub>4</sub>, and the absorbance was measured at 450 nm using a microplate reader (Labsystems Multiscan MS). For analysis, comparably treated wells, but without addition of primary antibody, were used to tare measurements from antibody-treated wells.

**Protein Isolation from Tissues and Western Blot Analyses**—Tissues were dissected from adult C57BL/6J mice and immediately frozen in liquid nitrogen. For extraction the tissues (200 mg/experiment) were thawed on ice and weighed, and 5 volumes (ml/g of wet tissue) of chilled extraction buffer (TBS, 1% Nonidet P-40, 2 mM EDTA, and proteinase inhibitor mixture (Complete, Roche Applied Science)) were added before homogenization using a Polytron tissue homogenizer (Kinematica). Insoluble material was subsequently removed by centrifugation. Western blots of tissue extracts gave high background; therefore, immunoprecipitation was done with antibody generated in one species (e.g. guinea pig) followed by Western blotting the immunoprecipitate and detecting it with antibody generated in the other species (e.g. rabbit). For immunoprecipitation, clarified tissue extracts were preincubated with 50  $\mu$ l/ml gelatin-Sepharose resin for 1 h at 4 °C to remove proteins that nonspecifically bind to Sepharose. The tissue extracts were then incubated overnight with 5  $\mu$ g/ml polyclonal guinea pig anti-collagen XXIII ectodomain antibody covalently coupled to CNBr-activated Sepharose. Following the removal of unbound material, the immobilized antibody-antigen complexes were split, and one part was subjected to collagenase digestion (see below) whereas the other part was treated identically but without the addition of collagenase. Antibody-bound collagen XXIII was eluted from the Sepharose using the high pH elution method described above, and the eluates were concentrated in a SpeedVac concentrator (Savant). After reduction the immunoprecipitates were separated by 10% SDS-PAGE and subsequently blotted on nitrocellulose membranes using standard procedures. The concentrated collagen XXIII was detected using the rabbit polyclonal anti-collagen XXIII ectodomain antibody followed by incubation with an horseradish peroxidase-coupled swine anti-rabbit IgG (DakoCytomation). Lumi-

	<b>NC1</b>						<b>TM domain</b>					
Mu	MGAGERAAGG	GGGA	-----	-----	QDPGA	CGGSRALSAL	CLLLSVGSAA					39
Ra			T-			A	G					
Hu	P	-	DAGKGNAA	GGGGGGRSAT	TA		V					
Ch	---SK--	S	S	-----	-----	T-	T				<b>LTAV</b>	
Mu	<b>ACLLLGAQAA</b>	ALHGRVAALE	QERELLRHAG	PSGALAAWAE	THLERLLREK							89
Ra			R									
Hu	-	V	Q	E	R	P	D	P			-	
Ch	V	KT	E Q	L	-	-GAAGP	--L D	LQ-	PRV	Q	L	
Mu												
		<b>NC1</b>				<b>Col 1</b>						
Mu	LDGVAKLRTV	REAPSECICPP	GPPGRRGKP	GRRGDPGPPG	QSGRDGYPPG							139
Ra		I	A	V								
Hu												
Ch	GE L	L		D M					E	A		
Mu	LGLDGKPLP	GPKGKGTTPG	DFGPRGAQQQ	DGVAGPPGPP	GPPGARGPPG							189
Ra			A									
Hu	-----	-----	-----	D								
Ch	P		EA	V	D R	AT						
Mu	DTGKDGPRA	QGPAGPRGEP	GQDGEMGPKG	PPGPKGEPGI	PGKKGDDGMA							239
Ra		E	S				T				IP	
Hu			K				V				TP	
Ch		P	P LK	E	LV AR		M	Q			IP	
Mu	<b>NOPLGPPG</b>	PKGEPDVGVP	RGENGVDGIP	GLKGEPGHPG	VDGATGPRGP							289
Ra	S		Q	T	A		T	I				
Hu	S	---		SM	A	P	R	A	A			
Ch	--- E	LS		K ST L K	T A	P	EK	GV SI				
Mu	PGLKGEQGD	<b>VVIDYDGRILDALK</b>	GPPGPQ	GAPGPPGIPG	AKGELGLPGA							339
Ra												
Hu												
Ch	P			H	S	P	P	Q A	P		P	
Mu	PGIDGKPK	GPKGDPGEPG	PAGPKGETGE	MGLSGLPGAD	GPKGKGES							389
Ra												
Hu		Q		L	A				L			
Ch	P L	A	Q SA	IT		P				DAG		
Mu												
		<b>NC3</b>				<b>Col 3</b>						
Mu	SDHLQESLAQ	IIVEP	GPPGP	PGPPGPMGLQ	GIQGPKGLDG	AKGKGTGS--						437
Ra						A						
Hu	S		L			A						
Ch	PC	NN--H	P	L	M	P---				D	GE	
Mu	-GERGPHGLP	GPVGPPLIG	LPGTGKGEGR	PGEPGLDGFP	GPRGKGD	<b>S</b>						486
Ra												
Hu	S		Q						L		EK	
Ch	K	DA PA	LP A		K							
Mu	<b>ERGEKGERGV</b>	PGRKGVKQK	GEPGPPGLDQ	PCP	VGPDGLP	VPGCWHK	ter					533
Ra												
Hu												
Ch		I	A						A			

The reaction was stopped by adding EDTA to a final concentration of 20 mM.

**Immunohistochemistry**—Immunohistochemistry was performed on frozen OCT-embedded sections of embryonic mouse tissues preincubated in ice-cold methanol for 2 min, blocked for 1 h with 5% normal goat serum in phosphate-buffered saline containing 0.2% Tween 20, and incubated with the guinea pig collagen XXIII primary antibody overnight at 4 °C. This was followed by Cy<sup>TM</sup>3-conjugated goat anti-guinea pig IgG (DakoCytomation). Alexa 488-conjugated goat anti-rabbit IgG (Molecular Probes) was used as secondary antibody for co-staining rabbit collagen XXIII polyclonal antibody or detecting laminin 5 polyclonal antibody. Stained sections were analyzed, and pictures were taken with a confocal laser-scanning microscope (Leica TCS SL) using two lasers in parallel with the excitation wavelengths 488 nm for Alexa488 and 543 nm for Cy3.

**RESULTS**

**$\alpha 1$ (XXIII) Collagen cDNAs**—A BLAST search (18) of the dbEST (19) with triple helical sequences of FACIT collagens yielded a novel human EST clone (GenBank<sup>TM</sup> accession number W22262), which was purchased and sequenced. 5'-RACE extended the clone, and further data base searching revealed the 5'-end. The composite full-length human cDNA (GenBank<sup>TM</sup> accession number AY898961) encodes a unique collagen of 503 amino acid residues that strongly resembled type XIII collagen. It was assigned the next available number in the collagen family at the time,  $\alpha 1$ (XXIII).

A BLAST search with the human sequence identified a mouse IMAGE clone (accession number AW323570), which was sequenced and extended by RACE. The mouse  $\alpha 1$ (XXIII) cDNA sequence encodes 533 amino acid residues and was deposited in GenBank<sup>TM</sup> under the accession number AF410792. Degenerate nucleotide primer pairs, designed from stretches of identical amino acids in mouse and human, were used to amplify a chick corneal cDNA, yielding an ~100-bp product (15). BLAST searching of chicken ESTs yielded a partial *Gallus* cDNA (BU270831) (30), which was used to identify

**FIGURE 1. Amino acid sequence of the mouse (Mu), rat (Ra) (16), human (Hu), and chick (Ch)  $\alpha 1$ (XXIII) collagen chains aligned using the ExpASY Sim program.** Amino acids are shown in the rat, human, and chick polypeptides only where they differ from the mouse sequence. A hyphen indicates a gap where there is no residue in a particular chain to align with the others. The non-collagenous domains are highlighted in gray, and the transmembrane (TM) domain is shown in bold letters. The two glycine-substituted triplets are bold and highlighted in gray. The GenBank<sup>TM</sup> accession numbers are: mouse, AF410792; rat, AY158896; chick, AY876377; our human cDNA, AY898961; and previously published human cDNA, AY158895 (16).

nescence was visualized using the ECL Plus<sup>TM</sup> Western blotting detection reagents (GE Healthcare).

**Collagenase Digestion**—To confirm the specific detection of collagen XXIII in extracts from tissue homogenates, the antibody-precipitated protein was subjected to collagenase digestion as described previously (29). The incubation with 100 units/ml highly purified bacterial collagenase (CLSPA, Worthington Biochemicals) was carried out in 250  $\mu$ l of TBS containing 5 mM CaCl<sub>2</sub> and 1 mM 4-(2-aminoethyl)benzenesulfonyl fluoride (Roche Applied Science) for 4 h at 37 °C.

## $\alpha 1(\text{XXIII})$ Collagen

overlapping chicken genomic clones in the data base. The composite *in silico* derived cDNA was then amplified and sequenced from chick corneal mRNA, and the sequence was deposited in GenBank™ under accession number AY876377. The polypeptides encoded by the mouse, chick, and human cDNAs are shown in Fig. 1 along with the sequence of the rat (16). The cDNAs define three triple helical collagenous (Col) domains and four non-collagenous (NC) domains within the polypeptide numbered from the amino to the carboxyl end to make the designations consistent with the closest relatives in the collagen family (types XIII and XXV).

The mouse and chick  $\alpha 1(\text{XXIII})$  collagen NC1 domains are slightly shorter than the human 120-residue NC1 domain (Fig. 1). In all species, the amino termini do not have characteristics of signal peptides but instead have stretches of hydrophobic residues typical for membrane-spanning regions (31). The short amino-terminal cytoplasmic stretches of the orthologous molecules are quite different in length and amino acid composition. The Col 1 domain of mouse, rat, and the previously reported human (16)  $\alpha 1(\text{XXIII})$  collagen is 189 amino acid residues in length. In all the mammalian cDNAs, the Col 1 domain contains one Gly-X-Y triplet in which the glycine is exchanged by a serine or asparagine. The chicken collagen XXIII lacks this triplet and thus has a perfect Gly-X-Y triplet structure in Col 1 domain. The newly obtained human cDNA encodes a 150-residue instead of a 189-amino acid residue Col 1 domain. This difference may reflect alternative splicing of collagen XXIII mRNA, a feature prevalent in the related collagen XIII mRNA (32–36).

In human, mouse, and rat, the Col 2 domain is 75 residues, and the Col 3 domain is 111 residues in length. The chicken Col 2 domain contains one more triplet than the mammalian orthologs. A conserved feature in the Col 3 domains of all four species is a glycine-substituted triplet, SER (Ser-Glu-Arg). However, although all species contain at least one glycine-substituted triplet, there are no Gly-X-Gly-X-Y or Gly-X-Y-X-Y imperfections in any of the triple helical domains.

The NC2 domain is 14 amino acid residues in all species and is highly conserved. The 16-residue NC3 domain is conserved in mouse, rat, and human, but in the chick, where the domain is only 11 residues, conservation is moderate at best. Very high conservation, however, is again observed in the 18-residue carboxyl NC4 domain of all four species.

**Expression of Collagen XXIII mRNA in Immortalized Cell Lines**—Because collagen XXIII expression is up-regulated in rat metastatic prostate carcinoma cell lines (16), we tested whether collagen XXIII mRNA was expressed by various immortalized human cell lines. The first tested were SEG-1 (a Barrett-associated adenocarcinoma cell line), SKGT4 (an esophageal epithelial adenocarcinoma), A549 (a non-small cell lung adenocarcinoma cell line), and the SV40-transfected esophageal epithelial cell line HET-1A. All express the  $\alpha 1(\text{XXIII})$  mRNA (Fig. 2) although to a lesser extent than they express EMMPRIN mRNA, a molecule up-regulated in epithelium-derived tumor cells (37). The HET-1 and the SKGT4 cells expressed collagen XXIII mRNA to about the same degree. The lung adenocarcinoma expressed somewhat less, whereas the Barrett-associated adenocarcinoma had the highest expression. The finding that collagen XXIII is expressed by immortalized prostate cell lines

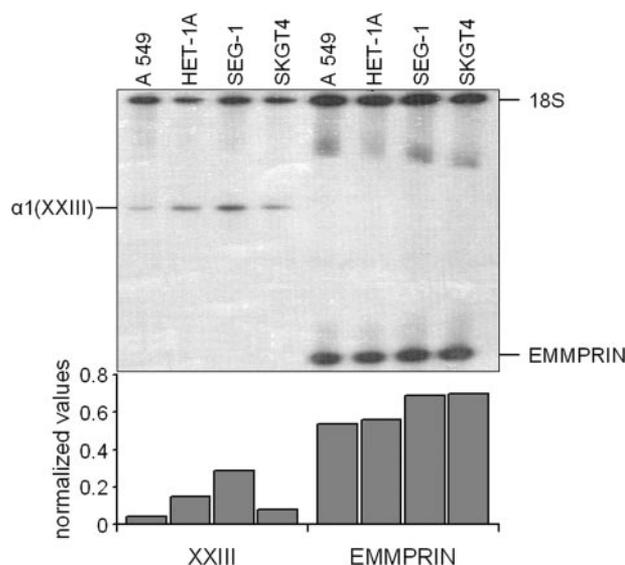
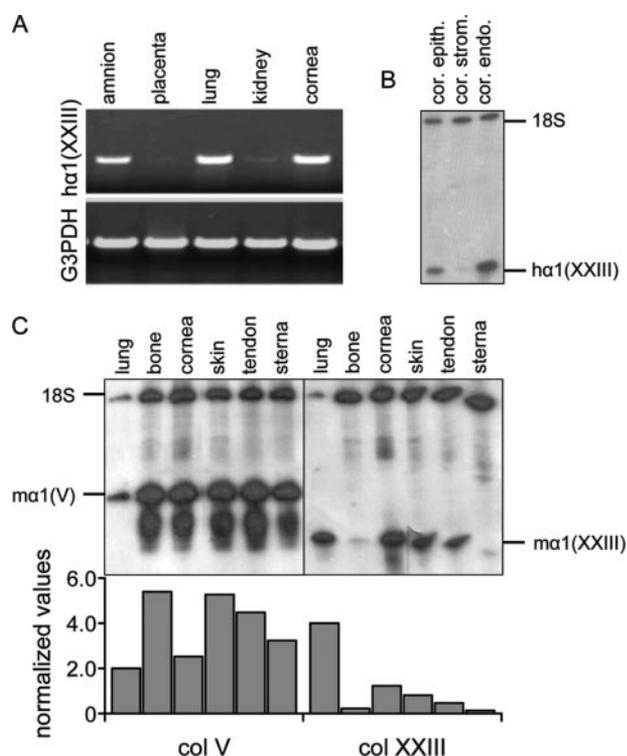


FIGURE 2. Expression of  $\alpha 1(\text{XXIII})$  collagen mRNA by relative RT-PCR in immortalized human cell lines A549 (a non-small cell lung adenocarcinoma line), HET-1A (an SV40-transfected esophageal epithelial line), SEG-1 (a Barrett-associated adenocarcinoma cell line), and SKGT4 (an esophageal epithelial adenocarcinoma). EMMPRIN expression was used as a control because it is expressed in most immortalized cells. The histogram shows the radioactive counts of the signals normalized in respect to the attenuated 18 S signals.

(16), the four immortalized cell lines tested here, and a fifth cell line, SV40 and adenovirus immortalized human corneal epithelial cells (see below), suggested that collagen XXIII might be widely expressed. This, however, was not the case for *in vivo* tissues.

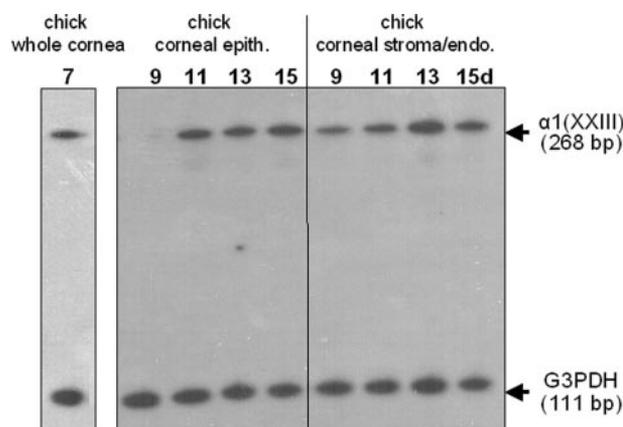
**Tissue Distribution of  $\alpha 1(\text{XXIII})$  Collagen mRNA**—RNA from a selection of human organs was surveyed to determine the spatial distribution of collagen XXIII. Strong signals for  $\alpha 1(\text{XXIII})$  collagen mRNA were seen in human amnion, lung, and cornea (Fig. 3A). Kidney and placenta yielded very minor signals. This suggests that collagen XXIII may be limited in distribution. Human donor corneal cDNA, prepared from RNA isolated from separated epithelial, stromal, and endothelial layers, showed that the corneal epithelium and endothelium contain significant amounts of  $\alpha 1(\text{XXIII})$  collagen mRNA, whereas the corneal keratocytes (*i.e.* differentiated stromal fibroblasts) synthesized almost none (Fig. 3B).

The distribution and abundance of  $\alpha 1(\text{XXIII})$  mRNA in mouse connective tissues and lung were analyzed. In initial experiments, the level of collagen XXIII mRNA was found to be very low compared with  $\alpha 1(\text{I})$  and  $\alpha 2(\text{I})$  collagen mRNAs (data not shown); therefore,  $\alpha 1(\text{XXIII})$  mRNA was compared with the minor fibrillar  $\alpha 1(\text{V})$  collagen chain. By relative semiquantitative RT-PCR (Fig. 3C), mouse lung contained the highest amount of the  $\alpha 1(\text{XXIII})$  collagen mRNA. The mRNA was also present in cornea, skin, and tendon but was hardly detectable in long bone (femur) and sternal cartilage. Taken together, these results suggest that  $\alpha 1(\text{XXIII})$  collagen mRNA is expressed by a limited number of organs, is a relatively non-abundant species, and, assuming comparable amplification efficiency for  $\alpha 1(\text{XXIII})$  and  $\alpha 1(\text{V})$ , is generally present at levels lower than  $\alpha 1(\text{V})$  collagen mRNA.



**FIGURE 3. Tissue distribution of human (A and B) and mouse (C)  $\alpha 1(\text{XXIII})$  collagen mRNA.** A, RT-PCR of  $\alpha 1(\text{XXIII})$  collagen mRNA in human amnion, placenta, lung, kidney, and cornea. B, human donor cornea (*cor.*), separated into its epithelial (*epith.*), stromal (*strom.*), and endothelial (*endo.*) cell layers. All PCRs were run in the linear range of amplification. In the non-radioactive PCR (A) glyceraldehyde-3-phosphate dehydrogenase (*G3PDH*) was used as the normalizer; in the radioactive relative RT-PCR (B) 18 S primers were used for normalization, attenuating the 18 S signal reproducibly with Ambion Competimers. C, tissue distribution of mouse collagen XXIII mRNA. RNAs from mouse lung, bone, cornea, skin, tendon, and sternal cartilage were subjected to relative RT-PCR for 29 cycles. Target primer pairs were used to detect mouse  $\alpha 1(\text{V})$  and  $\alpha 1(\text{XXIII})$  collagen mRNAs. The histogram shows the radioactive counts of the signals normalized in respect to the attenuated 18 S signals. *h*, human; *m*, mouse.

**Collagen XXIII mRNA Expression during Development**—The cornea achieves transparency through a defined developmental expression pattern that includes fibrillar and non-fibrillar collagens. The expression pattern of collagen XXIII mRNA was examined at important time points in chick corneal development. Whole 7-day cornea mRNA was assessed as was mRNA from separated corneal epithelial and stromal/endothelial layers at embryonic days 9, 11, 13, and 15. By day 7,  $\alpha 1(\text{XXIII})$  collagen mRNA is already present in the cornea as seen in Fig. 4. Because at day 9, expression is present in the stromal/endothelial layer but not in the epithelium, it is likely that the day 7 collagen XXIII mRNA signal is derived from the stromal/endothelial layer as well. And if the chick expression is similar to human expression, the 7- and 9-day signals may reflect mRNA predominantly from just the endothelium. By embryonic day 11, the pattern switches, and the amount of mRNA present in the epithelium is about the same as that of the stromal/endothelial layer. From this point on, the steady state level of  $\alpha 1(\text{XXIII})$  mRNA appears to fluctuate around a value equal to about one-half the amount of glyceraldehyde-3-phosphate dehydrogenase mRNA. The same  $\alpha 1(\text{XXIII})$  mRNA levels were also found at embryonic days 17 and 19 (data not shown).



**FIGURE 4. Collagen XXIII mRNA is expressed during development of the chick cornea.** Relative RT-PCR was used to assess the expression of chick  $\alpha 1(\text{XXIII})$  collagen mRNA with developmental time. Because the starting material was tissue poly(A)<sup>+</sup> RNA, glyceraldehyde-3-phosphate dehydrogenase (*G3PDH*) was used for normalization. For the 7-day (*d*) sample, whole corneal RNA was used. For embryonic days 9, 11, 13, and 15, the epithelium (*epith.*) was removed from the stromal/endothelial (*endo.*) layer with dispase and EDTA prior to RNA extraction and poly(A)<sup>+</sup> RNA isolation (25).

Whole mount and tissue section *in situ* hybridizations were used to examine expression of  $\alpha 1(\text{XXIII})$  collagen mRNA during mouse development. At E10.5 (Fig. 5A), hybridization is seen in the branchial arches and faintly in tail somites. At E11.5 (Fig. 5, B and C) the mRNA is present in tail somites, limbs, and slightly in brain. At E12.5 (Fig. 5D), in addition to the limb and tail hybridization, collagen XXIII mRNA signal is in the interdigital membrane where apoptosis will ultimately occur. Whisker follicles hybridize strongly. Branchial arch hybridization is, however, waning. On sections of tissue, hybridization in the E14.5 embryo shows collagen XXIII mRNA in the dura mater, brain, tongue, heart, and pharynx (Fig. 5E). Dura mater, umbilical cord, heart, gut, and lung signals are shown at higher magnification in Fig. 5, F–J. In addition, skin sections show that collagen XXIII mRNA is concentrated in the cells of the whisker (Fig. 5K) and hair (Fig. 5L) follicles.

**Tissue-specific Distribution of Collagen XXIII Protein**—Recombinantly expressed mouse collagen XXIII ectodomain was used to generate antibodies in rabbit and guinea pig. To assess the veracity of the recombinant protein, it was tested for collagenase sensitivity with highly purified bacterial collagenase (Fig. 6A). To monitor collagenase activity, the FLAG tag antibody was used on immunoblots of ectodomains treated and untreated with collagenase. The FLAG tag allows detection of the carboxyl-terminal NC4 domain produced by collagenase digestion because it prevents the small NC4 domain from running off the gel by adding  $\sim 10$  kDa in apparent molecular mass.

Antibodies were affinity-purified, and their specificity was assessed by cross-reactivity against the structurally related ectodomains of collagens XIII and XXV using solid phase assays (Fig. 6, B and C). To further exclude cross-reactivity against the His tag, which was used to purify the proteins, the His-tagged (but structurally unrelated) protein unc5h2 was included in the solid phase assays. The results indicate that the polyclonal antibody against collagen XXIII produced in guinea pig reached saturation at a dilution of 1:10,000 on collagen XXIII ectodomain and showed no cross-reactivity against the

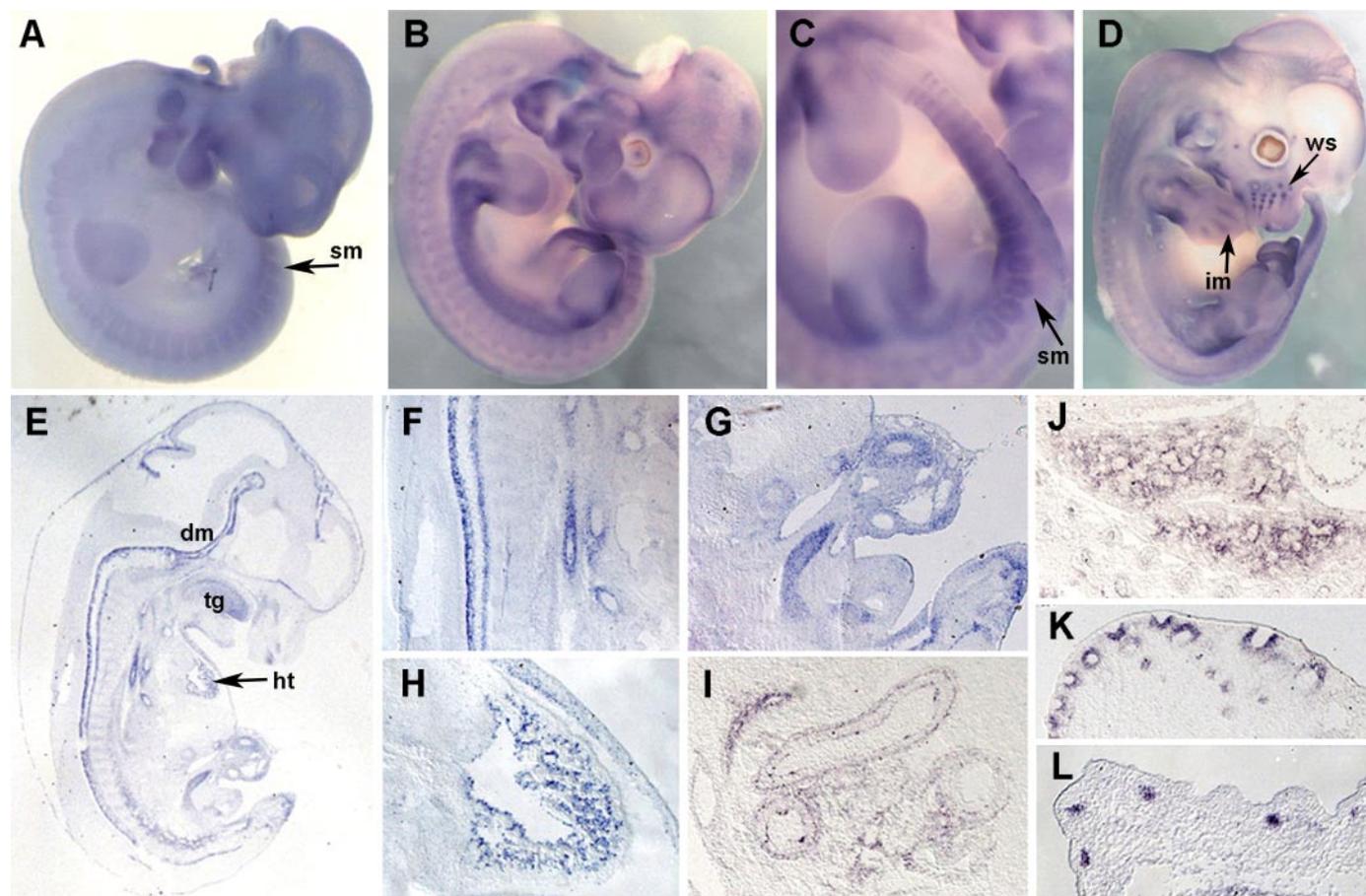


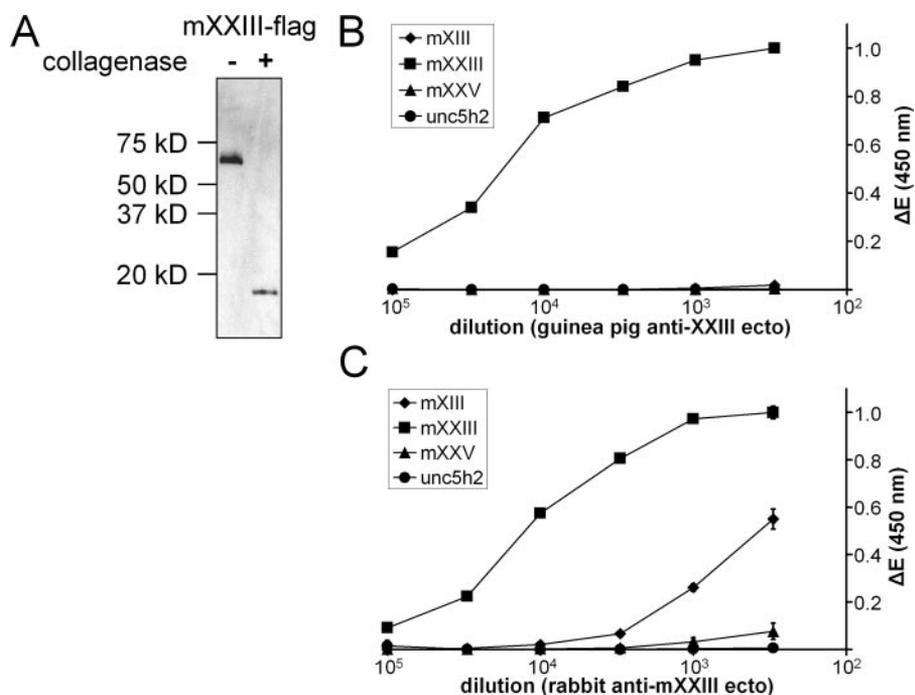
FIGURE 5. **Expression of collagen XXIII during mouse development.** A–E, whole mount *in situ* hybridization. A, hybridization is first seen faintly in E10.5 somites. B and C, by E11.5, collagen XXIII mRNA is seen in the somites and in limbs and faintly in brain. D, at E12.5, strong hybridization is also found in the interdigital membranes where apoptosis will occur and in the whisker follicles. E, an E14.5 whole embryo section where brain, dura mater, tongue, heart, and pharynx contain collagen XXIII mRNA. F, enlarged view of dura mater signal. G, umbilical cord. H, magnified heart signal. I, gut. J, lung. K, whisker follicle collagen XXIII signal. L, E15.5 skin section showing hybridization in hair follicles. sm, somites; ws, whisker follicle; im, interdigital membrane; ht, heart; tg, tongue; dm, dura mater.

other tested proteins (Fig. 6B). The polyclonal antibody produced in rabbit displayed a limited cross-reactivity against the ectodomain of collagen XIII, but at appropriate dilutions (1:10,000) its reactivity against collagen XXIII was about 100 times the cross-reactivity with collagen XIII (Fig. 6C).

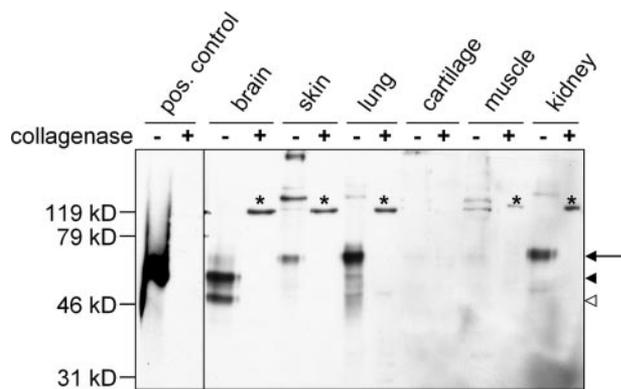
To analyze the tissue-specific distribution of collagen XXIII in adult mouse tissue, extracts of brain, skin, lung, cartilage, muscle, and kidney were immunoprecipitated with the guinea pig anti-mouse collagen XXIII antibody. The immunoprecipitated material was then Western blotted for reaction with the collagen XXIII antibody generated in rabbit. (This resulted in signals that were much cleaner than if tissue extracts were directly Western blotted.) As a positive control (Fig. 7), recombinant collagen XXIII was isolated from HEK 293 cells, which produce both the 68-kDa full-length form (arrow) and the 58-kDa shed form (black arrowhead). In mouse organs, skin and kidney contain almost exclusively the full-length form of the molecule. Lung contains mostly full-length form, but some shed material was detected as well (black arrowhead). Brain on the other hand contains almost no full-length transmembranous form. All appears to be processed to two forms: one is the expected size of the shed form, and the other is a lower molecular weight species that might represent additional processing

(open arrowhead). Cartilage contains no collagen XXIII, and muscle has little. In skin, lung, muscle, and kidney, a ~130-kDa band is seen that may represent non-reduced dimeric chains. In all tissues expressing collagen XXIII, the specific signal band was eliminated by digestion with highly purified bacterial collagenase. However, a higher molecular weight band appears (indicated by asterisks in Fig. 7). This likely represents a non-collagenous domain of collagen XXIII that is non-reducibly cross-linked to some other component. Such cross-linking might only become visible after collagenase digestion because, when not digested, the cross-linked product is too large to migrate into the gel.

Fig. 8 shows mouse E18.5 epithelial tissue cryosections (mouth facial skin, A–H; oropharynx and tongue, I–L; and small intestine, M–P) stained with each collagen XXIII antibody and with a laminin 5 antibody to assess the location of collagen XXIII in the tissues. By merging the signals from the guinea pig-derived and rabbit-derived antibodies (Fig. 8, C, G, K, and O) it is seen that both anti-collagen XXIII antibodies immunostain the same structures. The guinea pig antibody, however, reacts better with hair follicle cells than the one generated in rabbit (see Fig. 8C). Preincubation of the antibodies with excess recombinant collagen XXIII ectodomain abolished



**FIGURE 6. Sensitivity of recombinant collagen XXIII to collagenase treatment and specificity of antibodies.** A, recombinantly expressed FLAG-tagged collagen XXIII ectodomain was split into two aliquots for application to a 4–20% gradient SDS-polyacrylamide gel under reducing conditions. One aliquot was treated with highly purified bacterial collagenase prior to loading. B and C, specificity evaluation of collagen XXIII antibodies by solid phase assays. Polyclonal guinea pig (B) and rabbit (C) anti-collagen XXIII antibodies were tested in serial dilutions on microtiter plates coated with the ectodomains of collagens XIII, XXIII, and XXV or the unrelated His-tagged protein unc5h2. *m*, mouse; *ecto*, ectodomain.



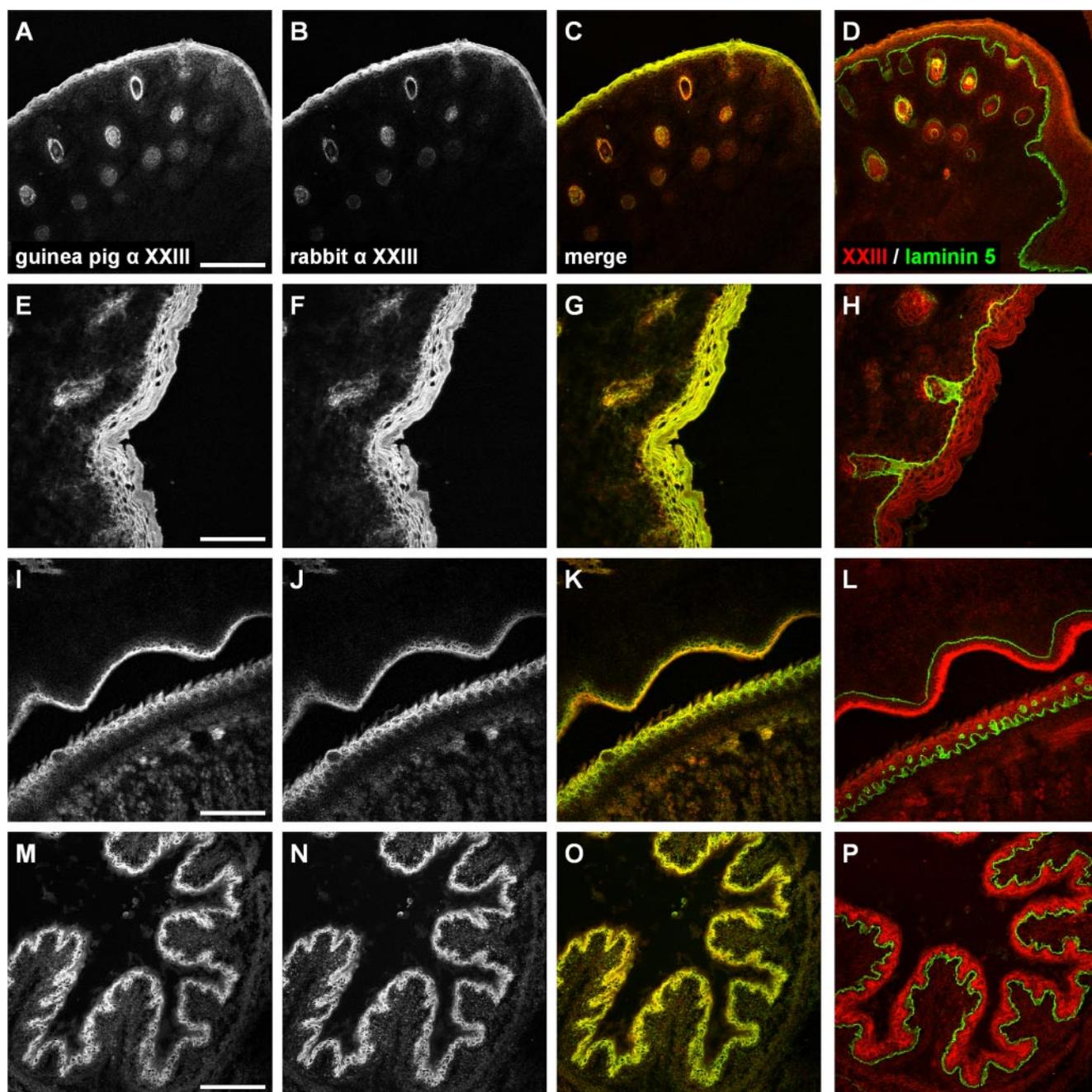
**FIGURE 7. Tissue distribution of collagen XXIII protein.** Collagen XXIII is detectable in Nonidet P-40 tissue extracts of brain, skin, lung, and kidney after immunoprecipitation with guinea pig antibody followed by Western blot detection of the immunoprecipitate with rabbit anti-collagen XXIII ectodomain polyclonal antibodies. In corresponding lanes loaded with collagenase-treated protein the signal is absent. The *arrow* indicates the molecular weight of the full-length protein, whereas in brain and lung there are also signals at the molecular weight of the ectodomain (*filled arrowhead*) and signals at even lower molecular mass (*open arrowhead*). Asterisks indicate a band that may represent a non-reducible cross-linking of an immunoreactive collagen XXIII NC domain with another protein. Conditioned medium of HEK 293-EBNA cells stably expressing the collagen XXIII ectodomain was used as a positive (*pos.*) control.

the signals (data not shown). In addition, co-staining of tissues for collagen XXIII and laminin 5, a basement membrane component, was performed, indicating that the collagen XXIII expression is restricted to the epithelial layer. The collagen likely interacts with the basement membrane, but it is not a component deposited into the basement membrane (Fig. 8, D,

H, L, and P). Also collagen XXIII is absent from the dermal layer except for the hair follicles (Fig. 8, A–H).

**Comparison of Human  $\alpha 1(\text{XXIII})$  with  $\alpha 1(\text{XIII})$  and  $\alpha 1(\text{XXV})$  Collagens**—The transmembrane collagens, type XXIII, type XIII, and type XXV, show a high degree of homology in their domain organization. They all contain three triple helical collagenous domains flanked by four non-collagenous domains with the transmembrane domain located in the NC1 domain. To examine the conservation between this subfamily, the amino acid sequences of human  $\alpha 1(\text{XXIII})$  collagen were aligned with human collagens  $\alpha 1(\text{XIII})$  and  $\alpha 1(\text{XXV})$ , using a best fit derived from alignments predicted by the ExpASY Sim program, coupled with a bias to align conserved exons derived from gene analyses (Fig. 9 and supplemental material). In the short cytoplasmic portion there is low conservation between the collagens XXIII, XIII, and XXV. Of the 26 amino acids that overlap in this domain in all chains, 4 amino acids are identical (15.4% identity), and an additional 4 residues are conservative changes (30.8% similarity). The transmembranous portions of NC1 domains are 26, 23, and 25 amino acid residues, respectively, for  $\alpha 1(\text{XXIII})$ ,  $\alpha 1(\text{XIII})$ , and  $\alpha 1(\text{XXV})$  collagens. 6 residues are identical across the chains (26%), and 2 other residues are conserved (34.8% similarity). In the 48 amino acids overlapping from the 60-, 60-, and 68-residue extracellular portions of the NC1 domains, 35.4% of the amino acid residues are identical, and 50% show similarity. The highest homologies can be detected around the putative furin cleavage site in the NC1 domain where 78% of the amino acid residues are identical. On the nucleotide level, this region contains a 30-bp stretch highly conserved between  $\alpha 1(\text{XIII})$  and  $\alpha 1(\text{XXV})$ .

Surprisingly by using the conserved exon structure bias, the triple helical domains of these three collagens do not align analogously. As suggested by the colors in the schematic depiction in Fig. 9, residues in the Col 1 domain of collagen XXIII are related to regions in both the Col 1 and Col 2 domains of the collagens XIII and XXV, whereas sequences from both the  $\alpha 1(\text{XXIII})$  Col 2 and 3 domains are analogous to residues within just the Col 3 domain of types XIII and XXV collagens. Consequently there is no equivalent of the collagen XIII/XXV NC2 domain in collagen XXIII. What is found instead is that the 14-residue  $\alpha 1(\text{XXIII})$  NC2 domain is most related to the 20- and 22-residue NC3 domains of  $\alpha 1(\text{XIII})$  and  $\alpha 1(\text{XXV})$  collagens (represented by *royal blue lines* in Fig. 9). The 11 residues that overlap between the NC2 domain of collagen XXIII and the NC3 domains of collagens XIII and XXV show 45.5% identity and 72.7% similarity. In addition, the collagen XXIII NC3 domain also resembles the NC3 domains of types XIII and



**FIGURE 8. Localization of collagen XXIII and basement membrane laminin 5 in several epithelia.** Immunohistochemistry was performed on frozen embryonic mouse sections (E18.5). Epithelia shown are facial skin (A, B, C, and D, and at higher magnification in E, F, G, and H), oropharynx and the dorsum of the tongue (I, J, K, and L), and the small intestine (M, N, O, and P). Sections were incubated with guinea pig anti-collagen XXIII antibody (A, E, I, and M), rabbit anti-collagen XXIII antibody (B, F, J, and N), and rabbit anti-laminin 5 polyclonal antibody (D, H, L, and P). For color reaction, primary antibody detection was followed by incubation with Cy3-conjugated goat anti-guinea pig (red) and Alexa488-conjugated goat anti-rabbit IgG (green). Note that guinea pig anti-collagen XXIII (A, E, I, and M) and rabbit anti-collagen XXIII (B, F, J, and N) polyclonal antibodies stain the same structures as indicated in the overlay pictures (C, G, K, and O; guinea pig  $\alpha$ XXIII, red; rabbit  $\alpha$ XXIII, green; co-localization, yellow). Co-staining with laminin 5 (D, H, L, and P; green) indicates that collagen XXIII (red) is concentrated in the epidermal layer and hair follicles (D and H) and other epithelial structures as shown for the oropharynx and the dorsum of the tongue (L) and the small intestine (P). The bars are 200  $\mu\text{m}$  in A–D, I–L, and M–P and 100  $\mu\text{m}$  in E–H.

XXV. Thus, within collagen XXIII, the NC2 and NC3 domains are related (discussed later).

As for the triple helical domains, the amino-most 54 residues of the  $\alpha 1(\text{XXIII})$  Col 1 domain have similarity with the Col 1 domains of collagens XIII and XXV (see *yellow-green* in Fig. 9). The 33 residues that overlap in all 3 chains, display 48.5% identity and 60.6% similarity. The carboxyl-most 132 residues of the  $\alpha 1(\text{XXIII})$  Col 1 domain are related to the Col 2 domains of

collagens XIII and XXV (colored *light blue* in Fig. 9). Here 107 residues overlap in the three chains with 53.3% identity and 57% conservation of amino acids. The Col 2 domain of type XXIII collagen has similarity with the amino-most 102 residues of the  $\alpha 1(\text{XIII})$  Col 3 domain and with the amino-most 78 residues of the  $\alpha 1(\text{XXV})$  Col 3 domain (*pink*). Of the overlapping 69 amino acids 39 are identical (56.5%), and 44 are conserved (63.8%). The remainder of the collagen XIII and XXV Col 3 domains

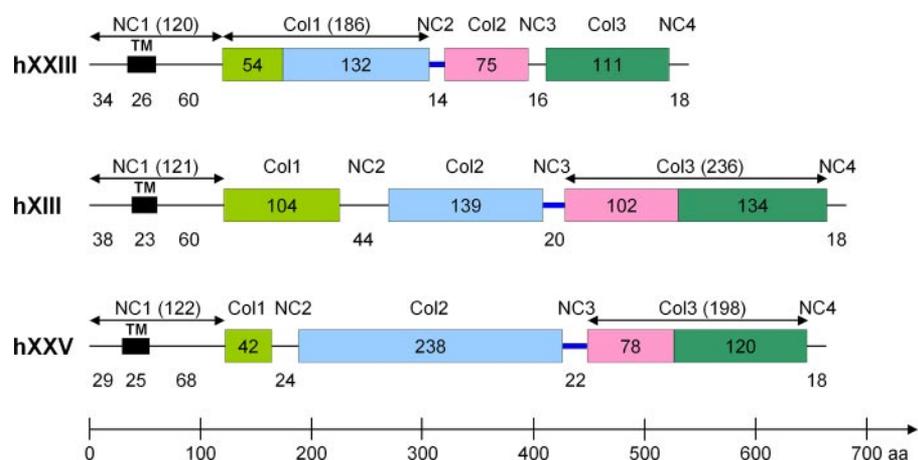


FIGURE 9. Comparison of collagens XXIII, XIII, and XXV. A schematic diagram of the three chains using color to indicate regions of similarity in the  $\alpha 1(\text{XXIII})$ ,  $\alpha 1(\text{XIII})$ , and  $\alpha 1(\text{XXV})$  chains is shown. Col 1 domain of  $\alpha 1(\text{XXIII})$  is similar to regions of both Col 1 and Col 2 domains of types XIII and XXV, whereas the Col 3 domains of types XIII and XXV are related to sequences in Col 2 and Col 3 domains of type XXIII collagen. In addition, the collagen XXIII NC2 domain is most related to the NC3 domains in types XIII and XXV. *TM*, transmembrane domain; *h*, human; *aa*, amino acids.



FIGURE 10. Both the NC2 and NC3 domains of human  $\alpha 1(\text{XXIII})$  collagen have coiled-coil character when aligned with the coiled-coil region in the NC3 domains of  $\alpha 1(\text{XIII})$  and  $\alpha 1(\text{XXV})$  collagens. The heptad repeats (*abc-defg*) indicated above the alignments correspond to the new alignment of collagen XIII and XXV NC3 domains with the NC2 domain of collagen XXIII (the top three lines); the *abcdefg* below the alignment indicates the heptads in the NC3 domain of  $\alpha 1(\text{XXIII})$  collagen. (Alignments are adapted from those presented in Refs. 38 and 39). Identical amino acids in NC domains are indicated by black boxes, and conserved residues are indicated by gray boxes.

resembles Col 3 domain of  $\alpha 1(\text{XIII})$  collagen (dark green in Fig. 9B). Among 90 overlapping residues in the three collagen types, there is 53.3% identity and 62.2% amino acid conservation. For the carboxyl-terminal NC4 domain, all chains have 18 amino acid residues with 61.1% identity and 72.2% similarity. The high conservation of the NC4 domain across chains of different collagen types might indicate a conservation of function for this part of the molecule.

## DISCUSSION

Collagen XXIII cDNAs have now been sequenced from four different species, three mammalian and one avian. The region in the collagen XXIII chain with the most species variability is the cytoplasmic domain followed by the latter half of Col 1 domain. Also the chick Col 1 domain lacks a glycine-substituted triplet that is present in the other species. Offsetting these differences are several notable regions of conservation in  $\alpha 1(\text{XXIII})$  collagen chains from the four species: the transmembranous region and Col 2, NC3, and Col 3 domains are highly conserved across the three mammalian species, and all four species show high conservation in their NC2 and NC4 domains.

The cDNAs from all species predict that the  $\alpha 1(\text{XXIII})$  collagen polypeptide lacks a typical signal sequence and is inserted in the plasma membrane in a type II orientation (this work and Ref. 16) as is true for the transmembranous collagen types XIII,

XVII, and XXV (1). Clearly by the number of triple helical and non-collagenous domains, within this transmembranous collagen subfamily collagen XXIII is not highly similar to collagen XVII but is related to collagens XIII and XXV. Therefore, we examined the homology more closely (Fig. 9). The highest conservation was detected in the Col 3 and NC4 domains, but local areas throughout the chains show similarity. Surprisingly the best alignment resulted when the Col 1 domain of type XXIII was aligned with both the Col 1 and Col 2 domains of types XIII and XXV and when both  $\alpha 1(\text{XXIII})$  Col 2 and Col 3 domains were aligned with just the Col 3 domain of the related colla-

gens. Thus, the collagen XXIII NC2 domain is not homologous with the NC2 domains of the collagens XIII and XXV but instead is homologous with their NC3 domains. This highlights a major difference among these members: collagen XXIII does not have a region analogous to the NC2 domains in collagens XIII and XXV. Because the number of triple helical domains is conserved in these three transmembranous family members, but the linear arrangement of comparable sequence is different between collagen XXIII versus collagens XIII and XXV, it suggests that after the collagen XXIII gene separated from a collagen XIII/XXV precursor gene, one or both genes experienced shuffling or insertions of exons to evolve into the current gene structures. And comparing the area around the furin cleavage site and the NC3 domains suggests that collagen XXIII may be closer to collagen XXV than to collagen XIII in local regions.

There is increasing evidence that coiled-coil domains within the non-collagenous NC domains of transmembrane collagens serve as oligomerization sites used to fold distinct areas of the forming molecules, thus facilitating trimerization. Hydrophobic residues in appropriate positions in heptad repeats have been reported in the NC1 and NC3 domains of collagens XIII, XXIII, and XXV (38, 39). These are proposed to locally aid in folding the triple helical domains, and in support of this idea, it has been shown that  $\alpha 1(\text{XIII})$  chains lacking the coiled-coil region in the NC1 domain incorrectly fold the Col 1 domain but not the Col 2 and 3 domains (38). These data support the suggestion that folding of Col 2 and 3 domains is assisted by the coiled-coil region in the NC3 domain. However, in these reports, the alignment of collagens XIII, XXIII, and XXV did not take into account the discontinuity shown in Fig. 9 from using conserved exons to dictate alignment, and thus, a coiled-coil region in the collagen XXIII NC2 domain was missed. This is shown in Fig. 10 by adding the collagen XXIII NC2 domain into the alignments of Latvanlehto *et al.* (38) and McAlinden *et al.* (39). This new comparison also indicates a heptad repeat conservation: by aligning the collagen XXIII NC2 domain with NC3 domain of  $\alpha 1(\text{XIII})$ , 6 identical (black boxed) and 4 conserved (gray boxed) residues are found in the region with 4

hydrophobic residues in "a" and "d" positions of the heptad repeats. Although the former comparison of the three NC3 domains also yielded 6 identical and 4 conserved residues in the NC stretch, it contained one less hydrophobic residue in an "a" position. Thus, by adjusting the polypeptide comparison of collagens XIII, XXIII, and XXV to consider a discontinuous similarity of the collagenous domains, a third potential coiled-coil region was revealed in a non-collagenous domain of collagen XXIII. Consequently every triple helical domain of  $\alpha 1$ (XXIII) collagen is preceded by an NC domain that can potentially participate in facilitating trimerization.

Collagen XXIII is up-regulated in rat prostate cancer cell lines with increasing metastatic potential (16). A total of five human immortalized cell lines were tested here, and all expressed the mRNA, suggesting that the molecule may be a marker for transformed cells in a fashion similar to what is observed for EMMPRIN (40–42). In contrast, the tissue distribution of collagen XXIII mRNA in normal human and mouse tissues is narrow. Collagen XXIII is most prevalent in lung followed by cornea, skin, tendon, and amnion. Minor amounts of mRNA were found in kidney and placenta. This distribution once again highlights the fact that many of the more recently discovered collagens are not most abundant in fibrous connective tissues but rather are found in organs where extracellular matrix is important but limited in quantity (for a review, see Ref. 43). We generated polyclonal antibodies against mouse collagen XXIII ectodomain in two species, guinea pig and rabbit. Immunofluorescence analysis revealed that collagen XXIII is a component of many epithelia, *i.e.* epidermis, epithelia of the small intestine, oropharynx, and the dorsum of the tongue. In addition hair follicles are positive for collagen XXIII. Co-staining with laminin 5 revealed that collagen XXIII is not a component of basement membranes but displays a cell surface staining pattern indicating that the full-length transmembranous form is prevalent in those tissues. Comparison of the immunofluorescence data (E18.5) with *in situ* hybridizations performed at earlier (E12.5–14.5) embryonic stages supports the finding that collagen XXIII is found in developing hair follicles and the epithelia of the pharynx and small intestine. In contrast, the *in situ* hybridizations show no signal in the epidermis. This implies that expression of the protein in the developing epidermis occurs at later developmental stages. On Western blots, the full-length polypeptide runs at an apparent molecular mass of 68 kDa, and the shed form runs at 58 kDa. Tissues such as skin appear to contain only the full-length form, whereas brain appears to have almost no full-length form at all. In addition to the shed collagen XXIII in brain, there is also a smaller band, which may represent further processing of collagen XXIII. This requires future investigation. Interestingly shedding is emerging as a common phenomenon in cell surface proteins believed to have a role in the interaction of cells with their extracellular matrix. It is not only a property of the transmembranous collagens (1, 17, 44) but also of molecules such as syndecan (45, 46) and EMMPRIN (47, 48).

Currently the function of collagen XXIII is unknown and can only be inferred. We do know that it is expressed not only in mature adult tissues but also in developing organs. The developmental expression of collagen XXIII in the chick cornea, in

concert with the data from the human cornea, suggests possible roles for the molecule. In the human cornea, the  $\alpha 1$ (XXIII) collagen mRNA is present in the epithelium and endothelium but only in trace quantities in the stroma. Thus, collagen XXIII is probably not involved in stromal transparency. The stromal/endothelial expression of chick  $\alpha 1$ (XXIII) collagen mRNA by day 7 is likely to be solely endothelial and suggests that collagen XXIII may be important in the production and/or assembly of Descemet membrane, which is visible by electron microscopy 2 days later. Moreover the later synthesis of the mRNA by the epithelium is consistent with the timing for establishing Bowman layer, a matrix that develops in the acellular zone under the epithelium around embryonic day 12 (49). Descemet membrane and Bowman layer are matrices that interface with the stromal matrix, and at these interfaces, the structural stability of the extracellular matrix is enhanced compared with the central stroma (50). Therefore, collagen XXIII may assist in enhancing the stability of the corneal interfacial matrix areas. The expression of collagen XXIII during mouse development also suggests a possible function for the molecule: collagen XXIII may have some involvement in the cascade of events that removes the interdigital membranes. The mRNA is expressed around the same time as the bone morphogenetic protein cascade that ultimately separates the digits.

In conclusion, we have presented the full-length sequences of mouse, chick, and human collagen XXIII polypeptides and have demonstrated that lung, cornea, and skin are major expressers of the  $\alpha 1$ (XXIII) collagen mRNA and protein. Polyclonal antibodies were generated, and they revealed the size of the full-length and shed forms of the polypeptide. Several types of cells are able to synthesize the molecule. Epithelial cells of several tissues make the mRNA and protein. It is likely that in lung, alveolar epithelial cells are also the source of the collagen. Because tendon is a source of the mRNA too, some types of fibroblasts express collagen XXIII as well. This is not universally true, however, because collagen XXIII mRNA is found in only trace amounts in human corneal fibroblasts and in mouse dermis. Because collagen XXIII is expressed on epithelial cell surfaces, it may play a role in the formation or maintenance of cell-cell contacts or in the polarization of epithelial cells.

*Acknowledgments*—We thank Fu-Shin Yu and K. Araki-Sasaki for immortalized human corneal epithelial cells; C. S. Yang, Luke Chen, Xiaochun Xu, David Beer, and David Schrumpp for the other human cell lines; and Sally Twining, Kim Vaughn, and Mitch Watsky for generously supplying human corneal epithelial, stromal, and endothelial cDNA.

## REFERENCES

1. Franzke, C. W., Bruckner, P., and Bruckner-Tuderman, L. (2005) *J. Biol. Chem.* **280**, 4005–4008
2. Pihlajaniemi, T., and Rehn, M. (1995) *Prog. Nucleic Acids Res. Mol. Biol.* **50**, 225–262
3. Pihlajaniemi, T., Myllylä, R., Seyer, J., Kurkinen, M., and Prockop, D. J. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84**, 940–944
4. Sund, M., Ylonen, R., Tuomisto, A., Sormunen, R., Tahkola, J., Kvist, A. P., Kontusaari, S., Autio-Harmainen, H., and Pihlajaniemi, T. (2001) *EMBO J.* **20**, 5153–5164
5. Kvist, A. P., Latvanlehto, A., Sund, M., Eklund, L., Vaisanen, T., Hagg, P.,

- Sormunen, R., Komulainen, J., Fassler, R., and Pihlajaniemi, T. (2001) *Am. J. Pathol.* **159**, 1581–1592
6. Nykvist, P., Tu, H., Ivaska, J., Kapyla, J., Pihlajaniemi, T., and Heino, J. (2000) *J. Biol. Chem.* **275**, 8255–8261
7. Giudice, G. J., Emery, D. J., and Diaz, L. A. (1992) *J. Investig. Dermatol.* **99**, 243–250
8. Diaz, L. A., Ratrie, H., III, Saunders, W. S., Futamura, S., Squiquera, H. L., Anhalt, G. J., and Giudice, G. J. (1990) *J. Clin. Investig.* **86**, 1088–1094
9. Jonkman, M. F., de Jong, M. C., Heeres, K., Pas, H. H., van der Meer, J. B., Owaribe, K., Martinez de Velasco, A. M., Niessen, C. M., and Sonnenberg, A. (1995) *J. Clin. Investig.* **95**, 1345–1352
10. Fujikawa, L. S., Foster, C. S., Gipson, I. K., and Colvin, R. B. (1984) *J. Cell Biol.* **98**, 128–138
11. Gordon, M. K., Fitch, J. M., Foley, J. W., Gerecke, D. R., Linsenmayer, C., Birk, D. E., and Linsenmayer, T. F. (1997) *Investig. Ophthalmol. Vis. Sci.* **38**, 153–166
12. Gipson, I. K., Spurr-Michaud, S., Tisdale, A., and Keough, M. (1989) *Investig. Ophthalmol. Vis. Sci.* **30**, 425–434
13. McGrath, J. A., Gatalica, B., Christiano, A. M., Li, K., Owaribe, K., McMillan, J. R., Eady, R. A., and Uitto, J. (1995) *Nat. Genet.* **11**, 83–86
14. Tasanen, K., Floeth, M., Schumann, H., and Bruckner-Tuderman, L. (2000) *J. Investig. Dermatol.* **115**, 207–212
15. Gordon, M. K., Gerecke, D. R., Hahn, R. A., Bhatt, P., Goyal, M., and Koch, M. (2000) *Investig. Ophthalmol. Vis. Sci.* **41**, S752
16. Banyard, J., Bao, L., and Zetter, B. R. (2003) *J. Biol. Chem.* **278**, 20989–20994
17. Hashimoto, T., Wakabayashi, T., Watanabe, A., Kowa, H., Hosoda, R., Nakamura, A., Kanazawa, I., Arai, T., Takio, K., Mann, D. M., and Iwatsubo, T. (2002) *EMBO J.* **21**, 1524–1534
18. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) *J. Mol. Biol.* **215**, 403–410
19. Boguski, M. S., Lowe, T. M., and Tolstoshev, C. M. (1993) *Nat. Genet.* **4**, 332–333
20. Araki-Sasaki, K., Ohashi, Y., Sasabe, T., Hayashi, K., Watanabe, H., Tano, Y., and Handa, H. (1995) *Investig. Ophthalmol. Vis. Sci.* **36**, 614–621
21. Gordon, M. K., Foley, J. W., Linsenmayer, T. F., and Fitch, J. M. (1996) *Dev. Dyn.* **206**, 49–58
22. Koch, M., Olson, P. F., Albus, A., Jin, W., Hunter, D. D., Brunken, W. J., Burgeson, R. E., and Champlaud, M. F. (1999) *J. Cell Biol.* **145**, 605–618
23. Koch, M., Laub, F., Zhou, P., Hahn, R. A., Tanaka, S., Burgeson, R. E., Gerecke, D. R., Ramirez, F., and Gordon, M. K. (2003) *J. Biol. Chem.* **278**, 43236–43244
24. Koch, M., Foley, J. E., Hahn, R., Zhou, P., Burgeson, R. E., Gerecke, D. R., and Gordon, M. K. (2001) *J. Biol. Chem.* **276**, 23120–23126
25. Schwabe, G. C., Trepczik, B., Suring, K., Brieske, N., Tucker, A. S., Sharpe, P. T., Minami, Y., and Mundlos, S. (2004) *Dev. Dyn.* **229**, 400–410
26. Visel, A., Thaller, C., and Eichele, G. (2004) *Nucleic Acids Res.* **32**, D552–D556
27. Mason, I. J., Taylor, A., Williams, J. G., Sage, H., and Hogan, B. L. (1986) *EMBO J.* **5**, 1465–1472
28. Veit, G., Kobbe, B., Keene, D. R., Paulsson, M., Koch, M., and Wagener, R. (2006) *J. Biol. Chem.* **281**, 3494–3504
29. Koch, M., Schulze, J., Hansen, U., Ashwodt, T., Keene, D. R., Brunken, W. J., Burgeson, R. E., Bruckner, P., and Bruckner-Tuderman, L. (2004) *J. Biol. Chem.* **279**, 22514–22521
30. Boardman, P. E., Sanz-Ezquerro, J., Overton, I. M., Burt, D. W., Bosch, E., Fong, W. T., Tickle, C., Brown, W. R., Wilson, S. A., and Hubbard, S. J. (2002) *Curr. Biol.* **12**, 1965–1969
31. Sonnhammer, E. L., von Heijne, G., and Krogh, A. (1998) *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **6**, 175–182
32. Peltonen, S., Rehn, M., and Pihlajaniemi, T. (1997) *DNA Cell Biol.* **16**, 227–234
33. Tikka, L., Pihlajaniemi, T., Henttu, P., Prockop, D. J., and Tryggvason, K. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85**, 7491–7495
34. Juvonen, M., Pihlajaniemi, T., and Autio-Harminen, H. (1993) *Lab. Investig.* **69**, 541–551
35. Juvonen, M., Sandberg, M., and Pihlajaniemi, T. (1992) *J. Biol. Chem.* **267**, 24700–24707
36. Pihlajaniemi, T., and Tamminen, M. (1990) *J. Biol. Chem.* **265**, 16922–16928
37. Kataoka, H., DeCastro, R., Zucker, S., and Biswas, C. (1993) *Cancer Res.* **53**, 3154–3158
38. Latvanlehto, A., Snellman, A., Tu, H., and Pihlajaniemi, T. (2003) *J. Biol. Chem.* **278**, 37590–37599
39. McAlinden, A., Smith, T. A., Sandell, L. J., Ficheux, D., Parry, D. A., and Hulmes, D. J. (2003) *J. Biol. Chem.* **278**, 42200–42207
40. Menashi, S., Serova, M., Ma, L., Vignot, S., Mourah, S., and Calvo, F. (2003) *Cancer Res.* **63**, 7575–7580
41. Polette, M., Gilles, C., Marchand, V., Lorenzato, M., Toole, B., Tournier, J. M., Zucker, S., and Birembaut, P. (1997) *J. Histochem. Cytochem.* **45**, 703–709
42. Muraoka, K., Nabeshima, K., Murayama, T., Biswas, C., and Koono, M. (1993) *Int. J. Cancer* **55**, 19–26
43. Knight, D. P. (2000) *J. Cell Sci.* **113**, 4141–4142
44. Hirako, Y., Nishizawa, Y., Sitaru, C., Opitz, A., Marcus, K., Meyer, H. E., Butt, E., Owaribe, K., and Zillikens, D. (2003) *J. Investig. Dermatol.* **121**, 1554–1556
45. Fitzgerald, M. L., Wang, Z., Park, P. W., Murphy, G., and Bernfield, M. (2000) *J. Cell Biol.* **148**, 811–824
46. Subramanian, S. V., Fitzgerald, M. L., and Bernfield, M. (1997) *J. Biol. Chem.* **272**, 14713–14720
47. Sidhu, S. S., Mengistab, A. T., Tauscher, A. N., LaVail, J., and Basbaum, C. (2004) *Oncogene* **23**, 956–963
48. Haug, C., Lenz, C., Diaz, F., and Bachem, M. G. (2004) *Arterioscler. Thromb. Vasc. Biol.* **24**, 1823–1829
49. Hay, E. D., and Revel, J. P. (1969) *Monogr. Dev. Biol.* **1**, 1–144
50. Linsenmayer, T. F., Gibney, E., and Fitch, J. M. (1986) *J. Cell Biol.* **103**, 1587–1593