Cloning and sequence analysis of cDNAs encoding the $\alpha 1$, $\alpha 2$ and $\alpha 3$ chains of mouse collagen VI

Rui-Zhu ZHANG,* Te-Cheng PAN,† Rupert TIMPL‡ and Mon-Li CHU*†§

Departments of *Biochemistry and Molecular Biology, and †Dermatology, Jefferson Institute of Molecular Medicine, Thomas Jefferson University, 233 South 10th Street, Philadelphia, PA 19107, U.S.A., and ‡Max-Planck-Institut für Biochemie, D-8033 Martinsried, Germany

cDNA clones encoding the $\alpha 1$, $\alpha 2$ and $\alpha 3$ chains of mouse collagen VI have been isolated by screening cDNA libraries with the corresponding human probes. The composite cDNAs for the $\alpha 1$, $\alpha 2$, and $\alpha 3$ chains are 2.5, 1.6 and 2.9 kb in size respectively. The $\alpha 1$ and $\alpha 2$ cDNAs encode the C-terminal portions of the chains as well as the entire 3'-untranslated regions, while the $\alpha 3$ cDNAs encode a central segment of 959 amino acids flanking the triple-helical domain. The deduced amino acid sequences share

86–88% identity with the human counterparts and 67–73% identity with the chicken equivalents. Alignment of the deduced amino acid sequences of mouse, human and chicken collagens reveal that the key features of the protein, including the cysteine residues, imperfections in the Gly-Xaa-Xaa regions, Arg-Gly-Asp sequences and potential N-glycosylation sites, are mostly conserved.

INTRODUCTION

Collagen VI, a major component of a widely distributed, abundant subclass of microfibrils (reviewed in Timpl and Engel, 1987), is a heterotrimer consisting of three different chains, i.e. $\alpha 1$, $\alpha 2$ and $\alpha 3$. Full-length cDNAs encoding the three chains from both human and chicken have been isolated and characterized (Chu et al., 1988, 1989, 1990; Bonaldo and Colombatti, 1989; Bonaldo et al., 1989, 1990; Koller et al., 1989; Trüeb et al., 1989). Sequence analyses showed that each chain contains a short central collagenous domain (336 amino acids for the $\alpha 1$ and $\alpha 3$ chains, and 335 amino acids for the $\alpha 2$ chain) flanked by large N- and C-terminal globular domains. The latter are composed of repetitive subdomains of ~200 amino acids, which are also found in von Willebrand factor, complement components, extracellular matrix proteins and integrin receptors (reviewed in Colombatti and Bonaldo, 1991). The $\alpha 1$ and $\alpha 2$ chains are ~1000 amino acids long, each consisting of one such repetitive subdomain in the N-globular region and two repeats in the C-globular region. The $\alpha 3$ chain is ~3000 amino acids long due to the presence of nine additional repeats in its N-globular domain and three other unrelated subdomains in the C-globular region (Bonaldo et al., 1990; Chu et al., 1990). In addition, several $\alpha 2$ and $\alpha 3$ chain variants, resulting from differential splicing of the repetitive non-collagenous subdomains, were discovered by cDNA cloning (Doliana et al., 1990; Saitta et al., 1990; Stokes et al., 1991).

Analyses of the primary structures suggest that the three constituent chains of collagen VI arose from a common ancestral gene by two successive gene duplications (Stokes et al., 1991). An earlier duplication separated the $\alpha 3$ gene from the ancestral gene for the $\alpha 1$ and $\alpha 2$ chains, and a more recent one produced the $\alpha 1$ and $\alpha 2$ genes. This hypothesis is supported by the observation that the $\alpha 1$ and $\alpha 2$ genes are syntenic on human chromosome 21 (Weil et al., 1988) as well as on mouse chromosome 10 (MacDonald et al., 1991), while the $\alpha 3$ gene is located on human chromosome 2 (Weil et al., 1988) and mouse chromosome 1 (Schurr et al., 1990).

Immunoelectron-microscopic studies have localized collagen VI microfibrils close to cells, nerves, blood vessels and collagen

fibrils (Bruns et al., 1986; Keene et al., 1988). Biochemical studies demonstrated that purified collagen VI or recombinant fragments bound cells, collagen fibrils, heparin and hyaluronan (Wayner and Carter, 1987; Aumailley et al., 1989; Bonaldo et al., 1990;

Figure 1 cDNA clones encoding the mouse collagen VI $\alpha 1$ (a), $\alpha 2$ (b) and $\alpha 3$ (c) chains

In each panel, a partial restriction map of the composite cDNA (thick line) is shown. The restriction enzymes sites are: A, *Apa*I; B, *Bgl*II; C, *Clal*; E, *Eco*RV; H, *Hind*III; P, *Pst*I; S, *Sma*I; X, *Xba*I. pA denotes a poly(A) tail. The cDNA clones (thin lines) isolated from mouse libraries are aligned with the protein domains predicted from previously described full-length human cDNAs. N1, N2, C1 and C2 indicate the 200-amino-acid repeats in the N- and C-globular domains, and C3 represents an additional structure in the C-globular domain (Chu et al., 1989, 1990); TH indicates triple-helical domain.

§ To whom correspondence should be addressed.

The nucleotide sequence data reported will appear in the EMBL, GenBank and BJ Nucleotide Sequence Databases under accession numbers Z18271, Z18272 and Z18279.

See the legend of Figure 2 for further details.

TTCTCCTGAGGGAGCTTTCTACCAAAAGGCAGATTATCGATGCCATCAACAAAGTCGTC	60	GGAGAGACGGGAGACATCGGCCCATGGCTCTCCAGGCAGAGATGCCATCCCTGGCAGC	1500
F F L R D F S T K R Q I I D A I N K V V		G E T G D I G P M G L P G R D G I P G S	
TATAAAGGAGGAGGATGCCAACACACGGGTGGCATCGAGCACCTGCTCGGAATCAC	120	CCAGAGACCTCGGAAGGACGGTGGCTCTGGCCGAAGGGCCCTGCAGGAGTAAAGGC	1560
Y K G G R H A N T R V G I E H L L R N H		P G D P G K D G G S G R R G P A G A K G	
TTGTGCGGAGGCTGGTAGCGCCTGGATGAGCGGGTCTCCAGATCGCCTTCGTGATC	180	AACAGAGTGGTCTGGCAGCAGCGCTCGAAGGCGAGCAGGTACAGAGGCTCACAG	1620
F V P E A G S R L D E R V P Q I A F V I		N R G G P G Q P G F E G E Q G T R G S Q	
ACCGAGGGAAGTCAGTGAGGATGCTCAGGATGTAGCCTGGCTTCAACCAGAAAGT	240	GGCCACCTGGTCCATTGGTCCCCAGGCTGATTGGGAACAAGNNNNCCGACCT	1680
T G G K S V E D A Q D V S L A L T Q K G		G P P G P I G P P G L I G E Q X X X G P	
GTCAGGTGTTCGCGTGGCGTGGAGAACATGACTGTAGGAGGTGGCAGATAGCC	300	CGAGAGGCGGAGGTACGGCTGGGGCTCTGGAGAACGTGGCAGAACTGGTCTCTGGG	1740
V K V F A V G V R N I D S E E V G K I A		R G G G G T A G A P G E R G R T G P L G	
TCCACAGTCCACAGCCTCCGGGTGGCAGCGTGCAGAACTGTCCGAATGAGTGAG	360	AGGAAGGTGAGCCAGGAGAGCCAGGCCGAAAGGAAGCATCGGAACCGTGACCCCGA	1800
S N S A T A F R V G S V Q E L S E L S E		R K G E P G E P G P K G S I G N R G P R	
ACAGTCTCGAGACTGTGATGATGCAATGATGAGACCCCTTGCCTGGTGTGACCGAT	420	GGAGAGCGGAGATGACGGGAGAGATGGGGTGGCAGTGAAGGACGACAGGCAAAAGA	1860
V S K A C N L E V I L G F D G S R D Q N		E T G D D G R D Q V G S E G R T G K K	
GTGTTGTGAGTCAGAGGGCCTTGAAGTGAAGTGAATCTTGAACAGAAATCAGC	480	GGAGAAAGAGGCTTCCCTGGGTATCGGGACCAAGGGCACCCCTGGTGAAGCCAGGGCA	1920
V F V F A V G V R N I D S E E V G K I A		G E R G F P G Y P G P K G T P G E P G A	
CAGATACAAAGGATCAGCTGAGTGGCAACAGCTGCCACATGTGAGGGTATCCGTGATG	540	GATGGACCAAGGACCAAGGCATCAGAGGTGCAAGGGAAATTCAGGACCTCCAGGG	1980
Q I Q R I S C S G N Q L P T V R V S V H		G P P P G P K G I R G R N G N S P P G	
GCGAAGCAGCCCTCTGGGCGAGTGGCCCTTGAAGTGTGCGAGTATCAGCGGAGCTG	600	GCAACTGGACAAAGGGAGACCTGGTACCAGGACCATCTGGTCAAGGGCAACCGA	2040
A N T P E A G F D F A Q P E L		A T G Q K G D P G Y P G P S G H K G N L	
TTTGAGAAAGTTTCGCAACATGCGAGTGCAGGCCCATACGCTCTCACAGCTGACAGCTG	660	GGCGACTCTTGTGATCAATGGCCCTCATCCAGAGCATCAGAGCAAAATGCGCTTGTCC	2100
F E K F R N M R S Q R P Y V L T A D T L		G D S V D Q C A L I Q S I R D K C P C	
AAGTTGTACAGAAAGTTCAGGCGAGTGTGCGCTGATCTGTGAAGGTAGTCAATCAC	720	TACGGCCCTCGAGTGGCTCTGTACTCCCAACAGAACTGGCTTTGCCCTGGATACCTCT	2160
K L Y Q N K S S P D T V K V I H		Y G P L E C P V L P T E L A F A L D T S	
TTACCGATGGAGCAGATGGAGACATGGCTGACTATACAGAGCATCTGAGGAATCCGG	780	GAGGGGGTACTCAGGACACCTTCAGCCGATCGGAGAGGTGCTTCTGGGCAATGSGGC	2220
F T D G A D G D H A D L Y R A S E E L R		E G V T Q D T F S R M R E V L L G I V G	
CAAGAGGTCACCAAGCCCTGATCTGGTGGCTTGAAGCTGTGGCTAACCTGGAGCGG	840	GATCTGACCATTCAGAGAGCAACTCTCCGCGGGTGGCCGCTGGCTGTGGTCACTAT	2280
Q A G A Q A L I L V G L E R V A N L E R		D L T I A E S N C P R G A R V A V V T Y	
CTGATGATCTGGAGTTCGGCGGGGCTTCATGTATGACAGGCCCTGAGGCTCAACAG	900	AACAATGAGGTGACTACAGAGATCCGGTTTGTAGCTTCAAGAGAAAGTCTGCCCTTTA	2340
L M H L E F G R G F N Y D R P L R L N Q		N N E V T T E I R F A S S K K K S A L L	
CTTGACTTAGACTATGAATGGCAGAGCAGCTTGACAATATGCTGAGAAAGCTTGTGT	960	GACAGTATTCAGAACTCCAAAGTGGCCTTGACCTTAAGCAGCAGAGCCTGGAGACCGCA	2400
L D L D Y E L A E Q L D N I A E K A C C		D S I Q N L Q V A L T S K Q Q S L E T A	
GGGCTTCATCAAGTCTCTGAGAGAGGGGAGCAGAGGGCCATTGGCAGCATTTGGG	1020	ATGTCAATCTGGCCAGGAATACATTCAAACTGTGAGGAGTGGCTTCCGATGAGGAAA	2460
G V P C K C S G E R G D R G P I G S I G		H S F V A R N T T F K R V R S G F P M R K	
CCAAAGGGTATCTCCGGGGAAGATGGCTACCGAGGCTACCCGGGAGACGAAGCGACCT	1080	GTGCTGTCTTCTCAGCAACAAGCCACAGGGCATCCCGCAGCTCCGAGAGGGCGTA	2520
P K G I S G E D G Y R G Y P G D E G D P		V A V F F S N K P T R A S P Q L R E A V	
GGTGAAGGAGTCCACCTGGTGAATGCAACAGGTTTCCAGGGCTTCCAGGCTCCAGCAG	1140	CTCAAGCTTTCGATGCGGGGATCACACCTTGTCTTACAGCCAGGAGGACCGACAG	2580
G E R G P V G N T Q G F G C P G Q		L K L S D A G I T P L F L T S Q E D R Q	
AGGGAGTGAAGGGTTCTCGAGATTCACAGGAGAAAGGTTGAATGGAGAAATCGGC	1200	CTCAACAGCTTTGAGATCAATACACAGCAGTGGGACATGCCCTGGTCTTCCTGCA	2640
R G V K G S R G F P G E K G E L G E I G		I N A L Q I N N T A V G H A L V L P A	
TTGGATGGTGTGATGCTGAAGAGGAGCAAGGGGCTGCTGGTCTTCTTGGAGAAAAG	1260	AGGAGGATCTCAGAGCTTCTGGAAGAAATGCTTACCTGCCATGTGTGCTGGACATT	2700
L D G L G E E G D K G L G S S G E K		R R D L T D F L K N V L T C H V C L D I	
GGGAGTCTGGAAGAGGGGTGCAAAAGACCAAGGAGCAAAAGGCGAGAGGAGAT	1320	TGCAATTCAGCCATCTTGGCTTCGGCAGCTGGAGGCTTCTTCCAGGACCGGAGG	2760
G S P G R G D K G P K G D K G E R G D		C N I D P S C G F G S W R P S F R D R	
GTCGGGATTCAGAGTGACCGGGTGAATCAGGAGCGGACAGCAGAGGAGGACCAAA	1380	GCAGCAGGAGTGTGAGACATAGACCTGGCTTTCATCTGGAGAGCTCAGAGGCTACC	2820
V G I R G D P G D S G R D S Q Q R G P K		A A G S D V D I D L A F I L D S S E A T	
	1440	ACACTGTTCCAGTCAATGAGATGAAGAAATACATAGGCTATGTATGATCAGAGCTG	2877
		T L F Q F N E M K K Y I G Y V I R Q L	

Figure 4 Nucleotide and deduced amino acid sequence of the composite mouse $\alpha 3(\text{VI})$ collagen cDNA

See the legend of Figure 2 for further details.

McDevitt et al., 1991; Kielty et al., 1992; Specks et al., 1992). Together, these data suggest that collagen VI mediates cell-matrix interactions. The precise biological functions of this ubiquitously expressed matrix protein, however, are still largely unknown. With the recent advances in transgenic mouse and homologous recombination technologies, the mouse has become a frequently used model system for studying gene functions *in vivo*. As an initial step towards using this approach to study the function of collagen VI, we have isolated and characterized cDNA clones for the three chains of the mouse protein.

MATERIALS AND METHODS

Two different mouse cDNA libraries were used for the screening of clones. One was prepared from 5 μg of mouse fibroblast poly (A)⁺ RNA in the $\lambda\text{ZAP II}$ vector as described (Chu et al., 1990). The other, made from mRNAs of 19-day mouse embryos in the λgt11 vector, was obtained from Clontech Laboratories (Palo

Alto, CA U.S.A.). The libraries were screened with cDNA probes specific for human $\alpha 1$, $\alpha 2$ and $\alpha 3$ chains of collagen VI (P18, P1 and P24 respectively; Chu et al., 1988) under low-stringency conditions, i.e. $6 \times \text{SSC}$ ($1 \times \text{SSC}$ is 0.15 M NaCl and 0.015 M sodium citrate, pH 7.0) at 55 °C. The cDNA probes were labelled with ^{32}P by nick translation (Sambrook et al., 1989). The filters were washed in $0.5 \times \text{SSC}$ at 55 °C. Positive clones were plaque-purified. Clones from the $\lambda\text{ZAP II}$ library were excised from the phage vector by infecting with R408 helper phage as suggested by Stratagene (La Jolla, CA, U.S.A.). The excision process yielded cDNA clones in the pBluescript SK (–) vector. cDNA inserts from the λgt11 library were subcloned into the pBluescript KS (+) vector.

Double-stranded cDNA clones were sequenced in both directions by the dideoxynucleotide termination method (Sanger et al., 1977) using the Sequenase kit (United States Biochemical, Cleveland, OH, U.S.A.) and oligonucleotide primers derived from the vector or the available cDNA sequences. DNA

Figure 5 Comparison of the $\alpha 1(\text{VI})$ collagen protein sequences of mouse, human and chicken

The protein sequence predicted from the mouse cDNA clones is aligned with the corresponding human and chicken sequences. Amino acid identity is indicated by dashes (—), and gaps (*) are introduced to allow maximal alignment. The positions of all cysteine residues are indicated (+). Triple-helical domains are boxed. Imperfections in triple helical domains, RGD sequences and potential N-glycosylation sites are shaded.

sequences were analysed using the PC/GENE computer program from IntelliGenetics Inc. (Mountain View, CA, U.S.A.).

RESULTS AND DISCUSSION

Isolation of mouse cDNA clones

Initial screening of the mouse fibroblast cDNA library using cDNA probes for the human $\alpha 1$, $\alpha 2$ and $\alpha 3$ chains of collagen VI yielded four independent clones coding for the $\alpha 1$ chain, a single clone for the $\alpha 2$ chain, but no positive clone for the $\alpha 3$ chain. Identities of the clones were confirmed by comparing partial DNA sequences obtained from both ends of the cDNA inserts with those of the corresponding human cDNAs. In order to isolate a cDNA for the $\alpha 3$ chain, a 19-day mouse embryonic cDNA library was screened with the human $\alpha 3(\text{VI})$ cDNA probe and yielded two positive clones.

Sequence analysis

The nucleotide sequences of the cDNAs encoding the three chains were determined from these positive clones (Figure 1).

Figure 6 Comparison of the $\alpha 2(\text{VI})$ collagen protein sequences of mouse, human and chicken

See the legend of the Figure 5 for further details.

Overlapping $\alpha 1(\text{VI})$ cDNA clones, *mCol6a1-1* to *mCol6a1-4*, covered 2529 nt of the message, and clone *mCol6a1-4* contained a short poly(A) tail. The $\alpha 2(\text{VI})$ cDNA, *mCol6a2-1*, was 1580 nt long, and the two $\alpha 3(\text{VI})$ cDNAs, *mCol6a3-1* and *mCol6a3-2*, were 1665 and 1205 nt respectively. These two cDNAs did not overlap, but comparison with the human cDNA sequence suggested that the gap was only 6 bp.

The composite $\alpha 1$ cDNA encoded 150 amino acids of the central triple-helical domain, the entire C-globular domain (433 amino acids) and 766 nt of 3'-untranslated region (3'-UT). A polyadenylation signal (AATAAA) was found 14 bp upstream of the poly(A) tail (Figure 2). The $\alpha 2(\text{VI})$ cDNA encoded a large portion of the C-globular domain (371 amino acids) and 450 nt of the 3'-UT (Figure 3). The latter may represent the entire 3'-UT, as a poly(A) signal is found 12 bp upstream of the 3' end of the cDNA. We previously identified three human $\alpha 2$ cDNA variants that differed in the C-globular domain (Chu et al., 1989; Saitta et al., 1990). The mouse cDNA clone isolated here corresponded to the most abundant variant, the $\alpha 2\text{C2}$ form. The $\alpha 3$ cDNA clones contained regions coding for 347 amino acids of the N-globular domain, the entire triple-helical domain (with a 6 bp gap between the two cDNAs) and 276 amino acids of the C-globular domain (Figure 4).

Comparison with the corresponding human and chicken cDNAs

Alignment of the deduced amino acid sequences of the three mouse collagen VI chains with previously described human and chicken sequences showed a high degree of similarity in the coding regions (Figures 5–7). Between the human and mouse amino acid sequences there were 86–88 % identical residues plus an additional 5 % conservative substitutions for all three chains. A lower degree of sequence identity, 67–73 %, was found between mouse and chicken chains, consistent with divergence between mammals and avians. In general, interspecies conservation was higher in the triple-helical domains (> 90 %) compared with the

Figure 7 Comparison of the $\alpha 3(\text{VI})$ collagen protein sequences of mouse, human and chicken

See the legend of Figure 5 for further details.

N- or C-globular domains because of the invariant glycine residues in the triple-helical domains. The length of the globular domain was not conserved in these three species, as several insertions and deletions of one or more amino acids were required to allow maximal alignment of the sequences (Figures 5–7).

The key structural features of the protein predicted by the partial mouse cDNA clones are mostly conserved. These features include the cysteine residues, Arg-Gly-Asp sequences, imperfections in the triple helices, and potential N-glycosylation sites. In particular, all cysteine residues shown in the mouse sequences are conserved, as is the case for imperfections in the Gly-Xaa-Xaa sequences of the $\alpha 1$ and $\alpha 3$ chains (Figures 5–7). The results support our previous suggestions that both structural features serve important functions: the cysteine residues participate in the formation of triple-helical monomers as well as their further association into dimers and tetramers, and Gly-Xaa-Xaa imperfections introduce flexibility into the triple-helical domains, resulting in supertwisting of the monomer (Chu et al., 1988). It is noteworthy that two Gly-Xaa-Xaa imperfections are present in the triple-helical domains of both mouse and human $\alpha 2$ chains, but one of these two imperfections is absent in the chicken because of an alanine to glycine substitution (Chu et al., 1988; Trüeb et al., 1989; Constantinou and Jimenez, 1991). A similar substitution of a glycine in the chicken $\alpha 1$ chain (Chu et al., 1988; Bonaldo and Colombatti, 1989) results in a shorter second Gly-Xaa-Xaa imperfection in the triple-helical region compared with the corresponding regions of the human and mouse (four versus seven amino acids).

Alignment of the sequences showed that most of the RGD sequences are conserved in all three species, except that the chicken $\alpha 3$ chain lacked two of the five Arg-Gly-Asp sequences conserved between human and mouse (Figures 5–7). The inter-

species conservation supports the importance of these Arg-Gly-Asp sequences in cell-matrix interactions, as previously demonstrated by Aumailley et al. (1989).

In the regions compared, most putative N-glycosylation sites are conserved in human, mouse and chicken (Figures 5–7). The exceptions are: (1) the human $\alpha 3$ chain contains an extra Asn-Gly-Thr sequence in the triple-helical domain as compared with the mouse and chicken; and (2) the chicken $\alpha 1$ chain lacks two of the three sites conserved between human and mouse. The interspecies conservation is consistent with the previous finding that collagen VI is highly glycosylated. It has been shown that most of the potential glycosylation sites in human collagen VI are utilized, and these modifications may limit its ability to form lateral aggregates (Fujiwara et al. 1991).

The 3'-UTs of both the $\alpha 1$ and $\alpha 2$ genes show little interspecies similarity. It is of interest to note that the length of the 3'-UT of each gene varies greatly between these three species. For the $\alpha 1$ gene, the region is 980, 770 and 1400 bp in mouse, human and chicken respectively (Bonaldo et al., 1989; Chu et al., 1989). The 3'-UT of the $\alpha 2$ gene is 298, 467 and 973 nt in human, mouse and chicken respectively (Chu et al., 1989; Trüeb et al., 1989). This is in contrast to the situation found for several other members of the collagen gene family in that a higher degree of interspecies conservation in the 3'-UTs was noted (Myers et al., 1986). As a result of the differences in the length of the 3'-UTs, the mRNAs for the human $\alpha 1(\text{VI})$ and $\alpha 2(\text{VI})$ chains (4.2 and 3.5 kb respectively) are readily separable by gel electrophoresis, while the two corresponding mouse mRNAs cannot be resolved under the same conditions (results not shown). It is known that the 3'-UTs of eukaryotic mRNAs may play a role in regulating the translation efficiency or stability of the mRNAs. Whether the lack of apparent interspecies conservation in the 3'-UTs of

collagen VI implies the lack of a regulatory function for this region remains to be studied.

In summary, we have isolated cDNA clones for the $\alpha 1$, $\alpha 2$ and $\alpha 3$ chains of mouse collagen VI. Comparison of the coding sequences in human, mouse and chicken identified highly conserved regions, which will be good candidates for future study of the structure and function of collagen VI. The mouse cDNA clones isolated in this study will be valuable specific hybridization probes for additional studies of the temporal and spatial expression of collagen VI transcripts using the mouse as a model system.

This work was supported by grants from the National Institutes of Health AR38912 and the Alexander von Humboldt Foundation. We thank Dr. John Gartland for his critical reading of the manuscript.

REFERENCES

- Aumailley, M., Mann, K., von der Mark, H. and Timpl, R. (1989) *Exp. Cell. Res.* **181**, 463–474.
- Bonaldo, P. and Colombatti, A. (1989) *J. Biol. Chem.* **264**, 20235–20239.
- Bonaldo, P., Russo, V., Bucciotti, F., Bressan, G. M. and Colombatti, A. (1989) *J. Biol. Chem.* **264**, 5575–5580.
- Bonaldo, P., Russo, V., Bucciotti, F., Doliana, R. and Colombatti, A. (1990) *Biochemistry* **29**, 1245–1254.
- Bruns, R. R., Press, W., Engvall, E., Timpl, R. and Gross, J. (1986) *J. Cell Biol.* **103**, 393–404.
- Chu, M.-L., Conway, D., Pan, T.-C., Baldwin, C., Mann, K., Deutzmann, R. and Timpl, R. (1988) *J. Biol. Chem.* **263**, 18601–18606.
- Chu, M.-L., Pan, T. C., Conway, D., Kuo, H.-J., Glanville, R. W., Timpl, R., Mann, K. and Deutzmann, R. (1989) *EMBO J.* **8**, 1939–1946.
- Chu, M.-L., Zhang, R.-Z., Pan, T.-C., Stokes, D., Conway, D., Kuo, H.-J., Glanville, R., Mayer, U., Mann, K., Deutzmann, R. and Timpl, R. (1990) *EMBO J.* **9**, 385–393.
- Colombatti, A. and Bonaldo, P. (1991) *Blood* **77**, 2305–2315.
- Constantinou, C. D. and Jimenez, S. A. (1991) *Matrix* **11**, 1–9.
- Doliana, R., Bonaldo, P. and Colombatti, A. (1990) *J. Cell Biol.* **111**, 2197–2205.
- Fujiwara, S., Shinkai, H. and Timpl, R. (1991) *Matrix* **11**, 307–312.
- Keene, D. R., Engvall, E. and Glanville, R., W. (1988) *J. Cell. Biol.* **107**, 1995–2006.
- Kielty, C. M., Whittaker, S., Grant, M. E. and Shuttleworth, C. A. (1992) *J. Cell Biol.* **119**, 979–990.
- Koller, E., Winterhalter, K. H. and Trueb, B. (1989) *EMBO J.* **8**, 1073–1077.
- MacDonald, G., Chu, M.-L. and Cox, D. R. (1991) *Genomics* **11**, 317–323.
- McDevitt, C. A., Marcelino, J. and Tucker, L. (1991) *FEBS Lett.* **294**, 167–170.
- Myers, J. C., Brinker, J. M., Kefalides, N. A., Rosenbloom, J., Wang, S.-Y. and Gudas, L. J. (1986) *Nucleic Acids Res.* **14**, 4499–4517.
- Saitta, B., Stokes, D. G., Vissing, H., Timpl, R. and Chu, M. L. (1990) *J. Biol. Chem.* **265**, 6473–6480.
- Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edn., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Sanger, F., Nicklen, S. and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467.
- Schurr, E., Skamene, E., Morgan, K., Chu, M.-L. and Gros, P. (1990) *Genomics* **8**, 477–486.
- Specks, U., Mayer, U., Nischt, R., Spissinger, T., Mann, K., Timpl, R., Engel, J. and Chu, M.-L. (1992) *EMBO J.* **11**, 4281–4290.
- Stokes, D. G., Saitta, B., Timpl, R. and Chu, M.-L. (1991) *J. Biol. Chem.* **266**, 8626–8633.
- Timpl, R. and Engel, J. (1987) in *Structure and Function of Collagen Types* (Mayne, R. and Burgeson, R. E., eds.), pp. 105–143, Academic Press, Orlando, FL.
- Trüeb, B., Schaefer-Wiemers, N., Schreier, T. and Winterhalter, K. H. (1989) *J. Biol. Chem.* **264**, 136–140.
- Wayner, E. A. and Carter, W. G. (1987) *J. Cell Biol.* **105**, 1873–1884.
- Weil, D., Mattei, M.-G., Passage, E., N'Guyen, V. C., Pribula-Conway, D., Mann, K., Deutzmann, R., Timpl, R. and Chu, M.-L. (1988) *Am. J. Hum. Genet.* **42**, 435–445.