The EMBO Journal vol.9 no.2 pp.385-393, 1990

Mosaic structure of globular domains in the human type VI collagen α3 chain: similarity to von Willebrand Factor, fibronectin, actin, salivary proteins and aprotinin type protease inhibitors

Mon-Li Chu, Rui-Zhu Zhang, Te-cheng Pan, David Stokes, Dorothy Conway, Huey-Ju Kuo[1], Robert Glanville[1], Ulrike Mayer[2], Karlheinz Mann[2], Rainer Deutzmann[2] and Rupert Timpl[2]

Departments of Biochemistry and Molecular Biology and Dermatology, Jefferson Institute of Molecular Medicine, Thomas Jefferson University, Philadelphia, PA 19107, [1]Shriners Hospital for Crippled Children, Portland, OR 97201, USA and [2]Max-Planck-Institut für Biochemie, D-8033 Martinsried, FRG

Communicated by R.Timpl

Human collagen α3(VI) chain mRNA (~10 kb) was cloned and shown by sequence analysis to encode a 25 residue signal peptide, a large N-terminal globule (1804 residues), a central triple helical segment (336 residues) and a C-terminal globule (803 residues). Some of the sequence was confirmed by Edman degradation of peptides. The N-terminal globular segment consists of nine consecutive 200 residue repeats (N1 to N9) showing internal homology and also significant identity (17–25%) to the A domains of von Willebrand Factor and similar domains present in some other proteins. Deletions were found in the N3 and N9 domains of several cDNA clones suggesting variation of these structures by alternative splicing. The C-terminal globule starts immediately after the triple helical segment with two domains C1 (184 residues) and C2 (248 residues) being similar to the N domains. They are followed by a proline rich, repetitive segment C3 of 122 residues, with similarity to some salivary proteins, and domain C4 (89 residues), which is similar to the type III repeats present in fibronectin and tenascin. The most C-terminal domain C5 (70 residues) shows 40–50% identity to a variety of serine protease inhibitors of the Kunitz type. The whole sequence contains 29 cysteines which are mainly clustered in short segments connecting domains N1, C1, C2 and the triple helix, and in the inhibitor domain. Five putative Arg-Gly-Asp cell-binding sequences are exclusively localized in the triple helical segment. The globular domains of human α1(VI) and α2(VI) chain [Chu *et al.* (1989) *EMBO J.*, 8, 1939–1946] are analogous to N1, C1 and C2 of α3(VI) chain, indicating that its extra domains have special functions in type VI collagen.

Key words: alternative splicing/microfibrillar collagen/polypeptide sequences/protease inhibitor/von Willebrand Factor

Introduction

Type VI collagen is a major component of a group of ubiquitously occurring tissue microfibrils and consists of three polypeptide chains, α1(VI), α2(VI) and α3(VI) (Timpl and Engel, 1987). As shown by electron microscopy (Furthmayr *et al.*, 1983; Jander *et al.*, 1984; von der Mark *et al.*, 1984) these chains are assembled into a short triple helix of 105 nm in length, flanked on each side by a large globular domain. This structure was confirmed by complete sequence analysis of human (Chu *et al.*, 1988, 1989) and chicken (Bonaldo *et al.*, 1989; Koller *et al.*, 1989) α1(VI) and α2(VI) chains each containing ~1000 amino acid residues with about one-third of them contributing to the triple helix. Both chains share ~35% identical residues (Chu *et al.*, 1989) and their genes are located in close proximity on human chromosome 21 (Weil *et al.*, 1988). Their globular structures consist of three 200-residue repeats with similarity to domains found in von Willebrand Factor, cartilage matrix protein, some integrins and complement components (Bonaldo *et al.*, 1989; Chu *et al.*, 1989; Koller *et al.*, 1989). This was interpreted to indicate homo- and heterotypic binding of these globular domains during self assembly and matrix formation.

The α3(VI) chain is apparently quite different from the α1(VI) and α2(VI) chains with molecular mass estimates ranging from 200 kd in tissues (Trüeb and Winterhalter, 1986) to 250–260 kd in cell cultures (Engvall *et al.*, 1986; Colombatti and Bonaldo, 1987; Colombatti *et al.*, 1987). The size of the triple helical sequences is virtually identical in all three chains but there are some structures contributed exclusively by the α3(VI) chain which are essential for oligomer formation (Chu *et al.*, 1988). In addition, the selective down-regulation of α3(VI) chain synthesis by γ-interferon was shown to be rate limiting for the secretion and matrix deposition of type VI collagen (Heckmann *et al.*, 1989). These data predict further unique structural elements in the α3(VI) chain not existing in the other type VI collagen chains. In the present study we have determined the whole sequence of the α3(VI) chain. The data show structures homologous to the α1(VI) and α2(VI) chains and 11 additional domains including some with similarity to von Willebrand Factor, fibronectin, the trypsin inhibitor aprotinin, and an unusual 122 residue sequence repeat.

Results

Isolation and characterization of cDNA clones

Three cDNA clones (P24, P108 and F376) (Figure 1), covering 3 kb of α3(VI) chain mRNA, have been isolated previously from a placenta and a fibroblast cDNA library and were partially sequenced (Chu *et al.*, 1988). Three strategies were used to extend the clone repertoire to nearly the full mRNA length. The insert of P108 was used to screen an oligo(dT) primed fibroblast cDNA library. Three clones (FM12, FM20 and FM22) with insert sizes of 3.5–3.8 kb were characterized and they all shared a 2.2 kb *Eco*RI fragment with a short poly(A)$^+$ tail at the 3′ end. The sequence upstream of this *Eco*RI site matched the sequence at the 3′ end of P108. Using a 0.4 kb *Eco*RI–*Hin*dIII

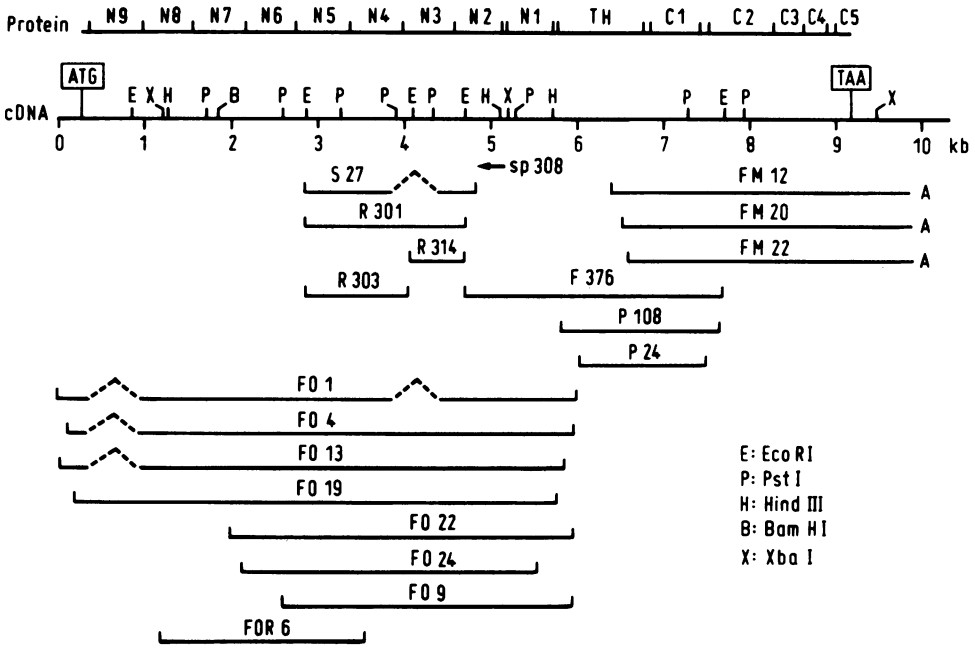

Fig. 1. Alignment and restriction maps of cDNA clones encoding ~9.9 kb of human α3(VI) chain mRNA. Potential splice sites in clones S27, F01, F04 and F013 are indicated by a dashed line. A denotes a poly(A) tail. SP 308 locates the position of a primer used to extend the library. On top the cDNA sequence is correlated with various globular domains (N1 to N9, C1 to C5) and the triple helical segment (TH) predicted for the α3(VI) chain (see Figure 7). ATG and TAA show the beginning and end of the open reading frame.

```
                                                                                                                CAGTTTGGAGCTCAGTCTTCCACCAAAGGC      30
CGTTCAGTTCTCCTGGGCTCCAGCCTCCTGCAAGGACTGCAAGAGTTTTCCTCCGCAGCTCTGAGTCTCCACTTTTTTGGTGGAGAAAGGCTGCAAAAAGAAAAAAGAGACGCAGTGAGTGGGAAAAGTATGCATCCTATTCAAACCTAAT     180
TGAATCGAGGAGCCCAGGGACACACGCCTTCAGGTTTGCTCAGGGGTTCATATTTGGTGCTTAGACAAATTCAAAATGAGGAAACATCGGCACTTGCCCTTAGTGGCCGTCTTTGCCTCTTTCTCTCAGGCTTTCCTACAACTCATGCC     330
                                                                                                  M  R  K  H  R  H  L  P  L  V  A  V  F  C  L  F  L  S  G  F  P  T  T  H  A   -1
CAGCAGCAGCAAGCAGCACAAGACTCTGCTGACATTATTTTCCTTATTGATGGATCAAACAACACCGGAAGTGTCAATTTCGCAGTCATTCTCGACTTCCTTGTAAATCTCCTTGAGAAACTCCCAATTGGAACTCAGCAGATCCGAGTG     480
Q  Q  Q  Q  A  A  Q  D  S  A  D  I  I  F  L  I  D  G  S  N  N  T  G  S  V  N  F  A  V  I  L  D  F  L  V  N  L  L  E  K  L  P  I  G  T  Q  Q  I  R  V      50
GGGGTGGTCCAGTTTAGCGATGAGCCCAGAACCATGTTTTCCTTGGACACCTACTCCACCAAGGCCCAGGTTCTGGGTGCAGTGAAAGCCCTCGGTTGCTGGTGGGAGTTGGCCAATATCGCTCGCCTTGATTTCGTGGTGGAGAACCAC     630
G  V  V  Q  F  S  D  E  P  R  T  M  F  S  L  D  T  Y  S  T  K  A  Q  V  L  G  A  V  K  A  L  G  C  W  W  E  L  A  N  I  A  R  L  D  F  V  V  E  N  H     100
TTCACCCGGGCAGGGGGCAGCCGCGTGGAGGAAGGGGTTCCCCAGGTGCTGGTCCTCATAAGTGCCGGGCCTTCTAGTGACGAGATTCGCTACGGGGTGGTAGCACTGAAGCAGGCTAGCGTGTTCTCATTCGGCCTTGGAGCCCAGGCC     780
F  T  R  A  G  G  S  R  V  E  E  G  V  P  Q  V  L  V  L  I  S  A  G  P  S  S  D  E  I  R  Y  G  V  V  A  L  K  Q  A  S  V  F  S  F  G  L  G  A  Q  A     150
GCCTCCAGGGCAGAGCTTCAGCACATAGCTACCGATGACAACTTGGTGTTTACTGTCCCGGAATTCCGTAGCTTTGGGGACCTCCAGGAGAAATTACTGCCGTACATTGTTGGCGTGGCCCAAAGGCACATTGTCTTGAAACCGCCAACC     930
A  S  R  A  E  L  Q  H  I  A  T  D  D  N  L  V  F  T  V  P  E  F  R  S  F  G  D  L  Q  E  K  L  L  P  Y  I  V  G  V  A  Q  R  H  I  V  L  K  P  P  T     200
ATTGTCACACAAGTCATTGAAGTCAACAAGAGAGACATAGTCTTCCTGGTGGATGGCTCATCTGCACTGGGACTGGCCAACTTCAATGCCATCCGAGACTTCATTGCTAAAGTCATCCAGAGGCTGGAAATCGGACAGGATCTTATCCAG     1080
I  V  T  Q  V  I  E  V  N  K  R  D  I  V  F  L  V  D  G  S  S  A  L  G  L  A  N  F  N  A  I  R  D  F  I  A  K  V  I  Q  R  L  E  I  G  Q  D  L  I  Q     250
GTGGCAGTGGCCCAGTATGCAGACACTGTGAGGCCTGAATTTTATTTCAATACCCATCCAACAAAAAGGGAAGTCATAACCGCTGTGCGGAAAATGAAGCCCCTGGACGGCTCGGCCCTGTACACGGGCTCTGCTCTAGACTTTGTTCGT     1230
V  A  V  A  Q  Y  A  D  T  V  R  P  E  F  Y  F  N  T  H  P  T  K  R  E  V  I  T  A  V  R  K  M  K  P  L  D  G  S  A  L  Y  T  G  S  A  L  D  F  V  R     300
AACAACCTATTCACGAGTTCAGCCGGCTACCGGGCTGCCGAGGGGATTCCTAAGCTTTTGGTGCTGATCACAGGTGGTAAGTCCCTAGATGAAATCAGCCAGCCTGCCCAGGAGCTGAAGAGAAGCAGCATAATGGCCTTTGCCATTGGG     1380
N  N  L  F  T  S  S  A  G  Y  R  A  A  E  G  I  P  K  L  L  V  L  I  T  G  G  K  S  L  D  E  I  S  Q  P  A  Q  E  L  K  R  S  S  I  M  A  F  A  I  G     350
AACAAGGGTGCCGATCAGGCTGAGCTGGAAGAGATCGCTTTCGACTCCTCCCTGGTGTTCATCCCAGCTGAGTTCCGAGCCGCCCCATTGCAAGGCATGCTGCCTGGCTTGCTGGCACCTCTCAGGACCCTCTCTGGAACCCCTGAAGTT     1530
N  K  G  A  D  Q  A  E  L  E  E  I  A  F  D  S  S  L  V  F  I  P  A  E  F  R  A  A  P  L  Q  G  M  L  P  G  L  L  A  P  L  R  T  L  S  G  T  P  E  V     400
CACTCAAACAAAAGAGATATCATCTTTCTTTTGGATGGATCAGCCAACGTTGGAAAAACCAATTTCCCTTATGTGCGCGACTTTGTAATGAACCTAGTTAACAGCCTTGATATTGGAAATGACAATATTCGTGTTGGTTTAGTGCAATTT     1680
H  S  N  K  R  D  I  I  F  L  L  D  G  S  A  N  V  G  K  T  N  F  P  Y  V  R  D  F  V  M  N  L  V  M  S  L  D  I  G  M  D  N  I  R  V  G  L  V  Q  F     450
AGTGACACTCCTGTAACGGAGTTCTCTTTAAACACATACCAGACCAAGTCAGATATCCTTGGTCATCTGAGGCAGCTGCAGCTCCAGGGAGGTTCGGGCCTGAACACAGGCTCAGCCCTAAGCTATGTCTATGCCAACCACTTCACGGAA     1830
S  D  T  P  V  T  E  F  S  L  N  T  Y  Q  T  K  S  D  I  L  G  H  L  R  Q  L  Q  L  Q  G  G  S  G  L  N  T  G  S  A  L  S  Y  V  Y  A  N  H  F  T  E     500
GCTGGCGGCAGCAGGATCCGTGAACACGTGCCGCAGCTCCTGCTTCTGCTCACAGCTGGGCAGTCTGAGGACTCCTATTTGCAAGCTGCCAACGCCTTGACACGCGCGGGCATCCTGACTTTTTGTGTGGGAGCTAGCCAGGCGAATAAG     1980
A  G  G  S  R  I  R  E  H  V  P  Q  L  L  L  L  L  T  A  G  Q  S  E  D  S  Y  L  Q  A  A  N  A  L  T  R  A  G  I  L  T  F  C  V  G  A  S  Q  A  N  K     550
GCAGAGCTTGAGCAGATTGCTTTTAACCCAAGCCTGGTGTATCTCATGGATGATTTCAGCTCCCTGCCAGCTTTGCCTCAGCAGCTGATTCAGCCCCTAACCACATATGTTAGTGGAGGTGTGGAGGAAGTACCACTCGCTCAGCCAGAG     2130
A  E  L  E  Q  I  A  F  N  P  S  L  V  Y  L  M  D  D  F  S  S  L  P  A  L  P  Q  Q  L  I  Q  P  L  T  T  Y  V  S  G  G  V  E  E  V  P  L  A  Q  P  E     600
AGCAAGCGAGACATTCTGTTCCTCTTTGACGGCTCAGCCAATCTTGTGGGCCAGTTCCCTGTTGTCCGTGACTTTCTCTACAAGATTATCGATGAGCTCAATGTGAAGCCAGAGGGGACCCGAATTGCGGTGGCTCAGTACAGCGATGAT     2280
S  K  R  D  I  L  F  L  F  D  G  S  A  N  L  V  G  Q  F  P  V  V  R  D  F  L  Y  K  I  I  D  E  L  N  V  K  P  E  G  T  R  I  A  V  A  Q  Y  S  D  D     650
GTCAAGGTGGAGTCCCGTTTGATGAGCACCAGAGTAAGCCTGAGATCCTGAATCTTGTGAAGAGAATGAAGATCAAGACGGGCAAAGCCCTCAACCTGGGCTACGCGCTGGACTATGCACAGAGGTACATTTTTGTGAAGTCTGCTGGC     2430
V  K  V  E  S  R  F  D  E  H  Q  S  K  P  E  I  L  N  L  V  K  R  M  K  I  K  T  G  K  A  L  N  L  G  Y  A  L  D  Y  A  Q  R  Y  I  F  V  K  S  A  G     700
AGCCGGATCGAGGATGGAGTGCTTCAGTTCCTGGTGCTGCTGGTCGCAGGAAGGTCATCTGACCGTGTGGATGGGCCAGCAAGTAACCTGAAGCAGAGTGGGGTTGTGCCTTTCATCTTCCAAGCCAAGAACGCAGACCCTGCTGAGTTA     2580
S  R  I  E  D  G  V  L  Q  F  L  V  L  L  V  A  G  R  S  S  D  R  V  D  G  P  A  S  N  L  K  Q  S  G  V  V  P  F  I  F  Q  A  K  N  A  D  P  A  E  L     750
GAGCAGATCGTGCTGTCTCCAGCGTTTATCCTGGCTGCAGAGTCGCTTCCCAAGATTGGAGATCTTCATCCACAGATAGTGAATCTCTTAAAATCAGTGCACAACGGAGCACCAGCACCAGTTTCAGGTGAAAAGGACGTGGTGTTTCTG     2730
E  Q  I  V  L  S  P  A  F  I  L  A  A  E  S  L  P  K  I  G  D  L  H  P  Q  I  V  N  L  L  K  S  V  H  N  G  A  P  A  P  V  S  G  E  K  D  V  V  F  L     800
CTTGATGGCTCTGAGGGCGTCAGGAGCGGCTTCCCTCTGTTGAAAGAGTTTGTCCAGAGAGTGGTGGAAAGCCTGGATGTGGGCCAGGACCGGGTCCGCGTGGCCGTGGTGCAGTACAGCGACCGGACCAGGCCCCGAGTTCTACCTGAAT     2880
L  D  G  S  E  G  V  R  S  G  F  P  L  L  K  E  F  V  Q  R  V  V  E  S  L  D  V  G  Q  D  R  V  R  V  A  V  V  Q  Y  S  D  R  T  R  P  E  F  Y  L  N     850
TCATACATGAACAAGCAGGACGTCGTCAACGCTGTCCGCCAGCTGACCCTGCTGGGAGGGCCGACCCCCAACACCGGGGGCCGCCCTGGAGTTTGTCCTGAGGAACATCCTGGTCAGCTCTGCGGGAAGCAGGATAACAGAAGGTGTGCCC     3030
S  Y  M  N  K  Q  D  V  V  N  A  V  R  Q  L  T  L  L  G  G  P  T  P  N  T  G  A  A  L  E  F  V  L  R  N  I  L  V  S  S  A  G  S  R  I  T  E  G  V  P     900
CAGCTGCTGATCGTCCTCACGGCCGACAGGTCTGGGGATGATGTGCGGAACCCCTCCGTGGTCGTGAAGAGGGGTGGGGCTGTGCCCATTGGCATTGGCATCGGGAACGCTGACATCACAGAGATGCAGACCATCTCCTTCATCCCGGAC     3180
Q  L  L  I  V  L  T  A  D  R  S  G  D  D  V  R  N  P  S  V  V  V  K  R  G  G  A  V  P  I  G  I  G  I  G  N  A  D  I  T  E  M  Q  T  I  S  F  I  P  D     950
TTTGCCGTGGCCATTCCCACCTTTCGCCAGCTGGGGACCGTCCAACAGGTCATCTCTGAGAGGGTGACCCAGCTCACCCGCGAGGAGCTGAGCAGGCTGCAGCCGGTGTTGCAGCCTCTACCGAGCCCAGGTGTTGGTGGCAAGAGGGAC     3330
F  A  V  A  I  P  T  F  R  Q  L  G  T  V  Q  Q  V  I  S  E  R  V  T  Q  L  T  R  E  E  L  S  R  L  Q  P  V  L  Q  P  L  P  S  P  G  V  G  G  K  R  D    1000
GTGGTCTTTCTCATCGATGGGTCCCAAAGTGCCGGGCCTGAGTTCCAGTACGTTCGCACCCTCATAGAGAGGCTGGTTGACTACCTGGACGTGGGCTTTGACACCACCCGGGTGGCTGTCATCCAGTTCAGCGATGACCCCAAGGCGGAG     3480
V  V  F  L  I  D  G  S  Q  S  A  G  P  E  F  Q  Y  V  R  T  L  I  E  R  L  V  D  Y  L  D  V  G  F  D  T  T  R  V  A  V  I  Q  F  S  D  D  P  K  A  E    1050
TTCCTGCTGAACGCCCATTCCAGCAAGGATGAAGTGCAGAACGCGGTGCAGCGGCTGAGGCCCAAGGGAGGGCGGCAGATCAACGTGGGCAATGCCCTGGAGTACGTGTCCAGGAACATCTTCAAGAGGCCCCTGGGGAGCCGCATTGAA     3630
F  L  L  N  A  H  S  S  K  D  E  V  Q  N  A  V  Q  R  L  R  P  K  G  G  R  Q  I  N  V  G  N  A  L  E  Y  V  S  R  N  I  F  K  R  P  L  G  S  R  I  E    1100
```

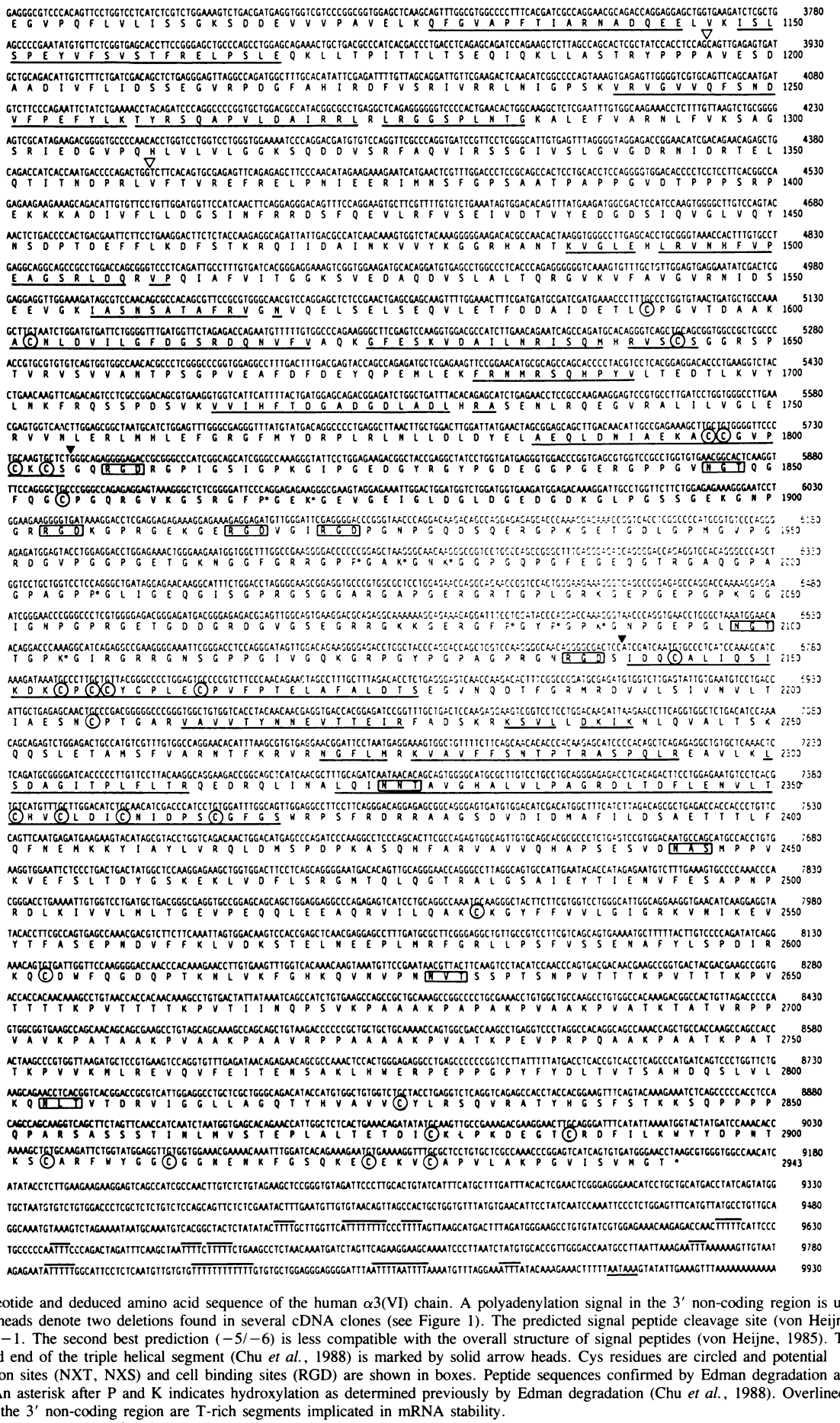

Fig. 2. Nucleotide and deduced amino acid sequence of the human α3(VI) chain. A polyadenylation signal in the 3′ non-coding region is underlined. Open arrow heads denote two deletions found in several cDNA clones (see Figure 1). The predicted signal peptide cleavage site (von Heijne, 1986) is numbered −1. The second best prediction (−5/−6) is less compatible with the overall structure of signal peptides (von Heijne, 1985). The beginning and end of the triple helical segment (Chu *et al.*, 1988) is marked by solid arrow heads. Cys residues are circled and potential N-glycosylation sites (NXT, NXS) and cell binding sites (RGD) are shown in boxes. Peptide sequences confirmed by Edman degradation are underlined. An asterisk after P and K indicates hydroxylation as determined previously by Edman degradation (Chu *et al.*, 1988). Overlined sequences in the 3′ non-coding region are T-rich segments implicated in mRNA stability.

fragment from the 5′ end of F376, we isolated from a primer extension library an overlapping 1.6 kb clone S27. Probing a random primed cDNA library with the insert of S27 yielded several more clones with insert sizes of either 0.6, 1.2 or 1.8 kb (R314, R303 or R301) (Figure 1). All these clones including S27 did not extend beyond an internal *Eco*RI site of α3(VI) mRNA (~2.9 kb from 5′ end) which apparently was not well protected by *Eco*RI methylase. In order to overcome this obstacle, we constructed another cDNA library using *Eco*RI adaptors instead of linkers, thereby eliminating the subsequent *Eco*RI digestion step. Screening of this library with a 0.4 kb *Eco*RI–*Pst*I fragment from the 5′ end of S27 yielded 30 cDNA clones (F0 series) (Figure 1) with inserts in the range 2.0–5.5 kb and containing several internal *Eco*RI sites. Together all these clones cover ~9.9 kb of α3(VI) mRNA (Figure 1).

Restriction enzyme mapping and/or sequence analysis of the various cDNA clones showed that some of them contained one or two deletions of 501 and 597 nt respectively (Figures 1 and 2). Since identical deletions were found in at least two independent clones they do not represent artifacts of the cDNA libraries but rather indicate alternative splicing of α3(VI) mRNA (see Discussion).

Nucleotide sequence of α3(VI) chain mRNA and deduced amino acid sequence

The various cDNA clones allowed us to determine a continuous sequence of 9930 nt with an open reading frame of 8904 nt (Figure 2). The open reading frame is preceeded at the 5′ end by a short non-coding region (255 nt) and is followed at the 3′ end by another 759 nt non-coding region with several stop codons in each reading frame and a short poly(A)$^+$ tail. A typical polyadenylation signal (AATAAA) starts 21 nt upstream of the poly(A) tail. The deduced amino acid sequence indicates a short signal peptide (25 residues), a large N-terminal (1804 residues) and C-terminal globule (803 residues), which are connected by a short triple helical segment (336 residues) characterized by repetitive Gly-X-Y triplets (Figure 2).

There is an apparent clustering of some functional amino acid residues in certain sections of the analyzed sequence (Figure 2). The 29 Cys residues are mainly restricted to the triple helical and C-terminal globular segments with some more found in the proximal and more distal portions of the N-terminal globule. A similar unequal distribution apparently exists for Asn residues present in nine N-linked glycosylation sites. The five putative RGD cell binding sites are exclusively localized within the triple helical segment.

An interesting feature of the 3′ non-coding region is its high AT content (72%) in the last 200 nt. In particular, long runs of Ts and a motif of one A followed by three or more Ts occur frequently in a region upstream of the poly(A)$^+$ tail (overlined in Figure 2). Similar sequence motifs have been found in the 3′ non-coding regions of many transiently expressed mRNAs, e.g. those coding for oncogenes, cytokines and lymphokines (Shaw and Kamen, 1986; Caput *et al.*, 1986) and were implicated as a recognition signal for the selective degradation of these mRNAs.

Sequence of peptides obtained from the globular domains of α3(VI) chain

About 50% of the triple helical sequence of the α3(VI) chain was previously supported by Edman degradation of peptides (Chu *et al.*, 1987, 1988) and has allowed the unequivocal identification of cDNA clones. Additional support was obtained from a globular segment NC1 (Odermatt *et al.*, 1983; Kuo *et al.*, 1989) which showed a characteristic HPLC profile of its tryptic peptides (Figure 3). Most of the major

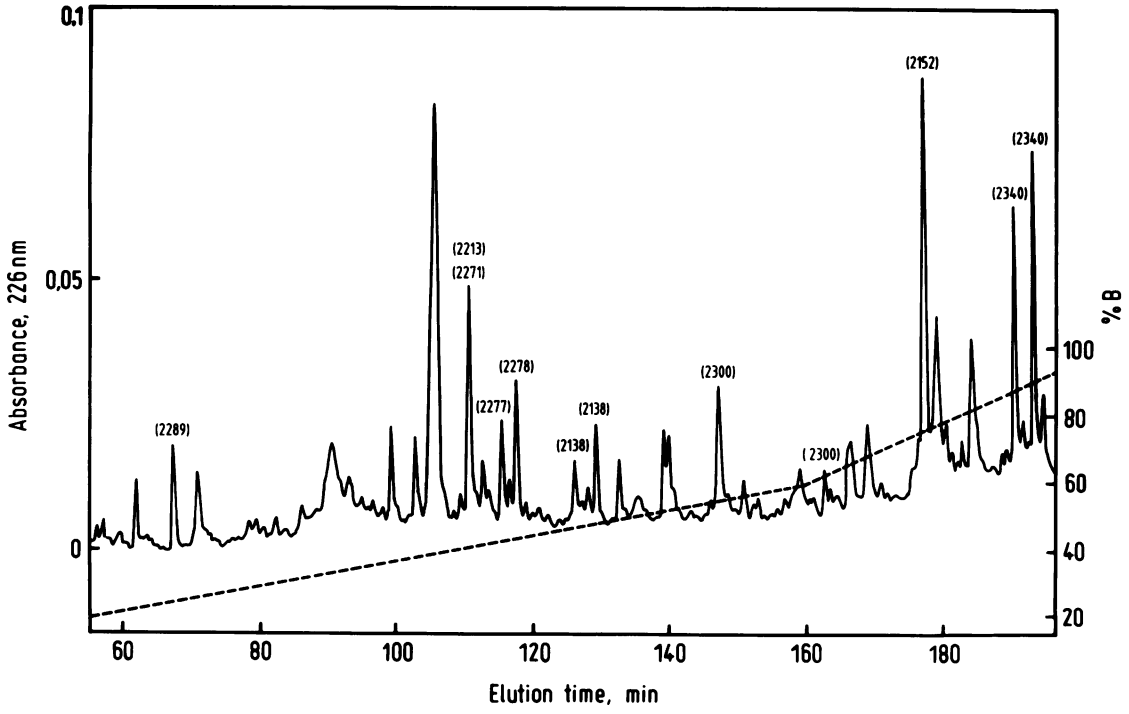

Fig. 3. HPLC separation of tryptic peptides obtained from fragment NC1 which corresponds to the C1 domain of α3(VI) chain. The starting positions (see Figure 2) of identified sequences are denoted on top of individual peaks. The digest was separated on a C_{18} column equilibrated in 0.1% trifluoroacetic acid which was eluted by an acetonitril (B) gradient (0–95%) as indicated by the dashed line.

peaks were sequenced and shown to fit exclusively to a section of the C-terminal globule within positions 2140–2365 (Figure 2) which is in agreement with the size estimate for this NC1 fragment. Sequence data for the N-terminal globule were obtained from 14 small peptides obtained from some other collagenase-resistant peptides with Lys-C protease and from pepsin solubilized collagen (Chu *et al.*, 1987, 1989). The peptide sequences fitted within positions 341–1804 of the sequence (Figure 2). Thus, ~20% of the cDNA sequence was confirmed for the globular domains available in the form of large fragments.

Comparison of amino acid sequences within distinct domains of α3(VI) chain

Sequence comparison of both globules for internal repeats and in comparison with other proteins revealed a number of interesting features, indicating that the globules of α3(VI) chain are organized in 14 different domains referred to as N9 to N1 and C1 to C5 (Figures 1 and 7). Domains N9 to N2 found in the N-terminal region are of rather uniform size (~200 residues) and show 31–40% identity (Figure 4). These domains show a distinct similarity to the three A domains of von Willebrand Factor (Titani *et al.*, 1986; Titani and Walsh, 1988) and two similar domains found in a cartilage matrix protein (Kiss *et al.*, 1989) which share 17–25% identical residues. As discussed previously (Chu *et al.*, 1989) some integrins and complement components possess similar domains.

Three more domains, N1, C1 and C2, which are more variable in size (184–248 residues) are located directly at the N- and C-terminal sites of the triple helix. They show a lower internal identity and identity to domain N2 of ~21% (alignment scores 2.1–5.8 SD; Table I) but are still of distinct similarity to N9 to N2. Their identities to the A domains of von Willebrand Factor and cartilage matrix protein are, however, rather low (11–14%). These three domains occupy positions in α3(VI) chain corresponding to the whole globular segments of the α1(VI) and α2(VI) chains (Bonaldo *et al.*, 1989; Chu *et al.*, 1989; Koller *et al.*, 1989). A comparison of these regions in all three chains (Table I) shows identity in the range 18–42% with a large variation of alignment scores (0.9–19.3 SD). The highest scores were observed in the comparison of analogous domains of α1(VI) and α2(VI) chains while the corresponding domains of α3(VI) were distinctly less similar. A particularly low and probably insignificant score (0.9–2.2 SD) was obtained for the N1 domains. Among the comparisons with other proteins a remarkable 32% identity was observed between a central

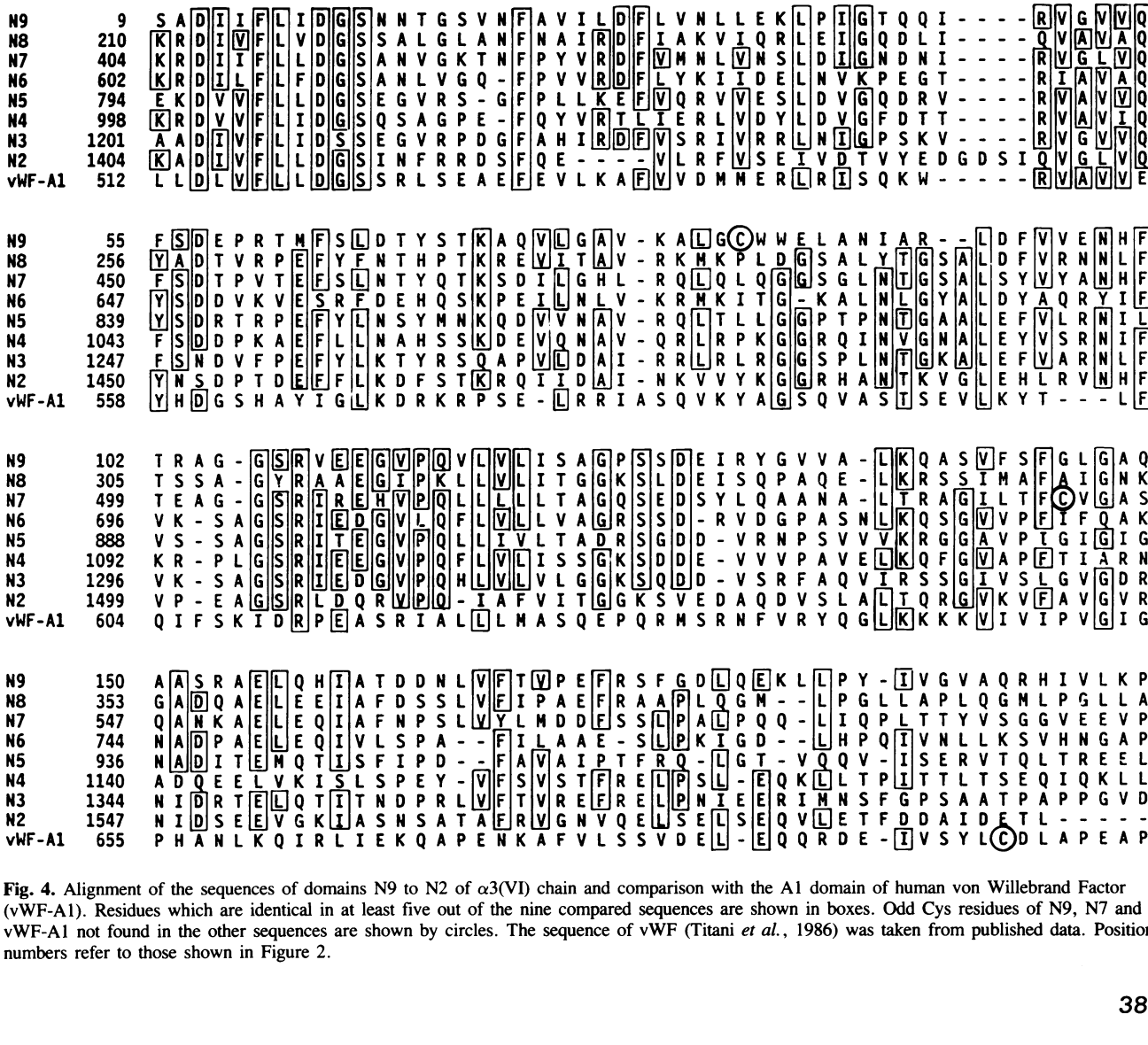

Fig. 4. Alignment of the sequences of domains N9 to N2 of α3(VI) chain and comparison with the A1 domain of human von Willebrand Factor (vWF-A1). Residues which are identical in at least five out of the nine compared sequences are shown in boxes. Odd Cys residues of N9, N7 and vWF-A1 not found in the other sequences are shown by circles. The sequence of vWF (Titani *et al.*, 1986) was taken from published data. Position numbers refer to those shown in Figure 2.

Table I. Sequence comparison of the 200 residue domains N2, N1, C1 and C2 of α3(VI) chain with each other and with homologous domains of α2(VI) and α1(VI) chain

% identity (alignment score) in comparison with									
Domain	α3(VI)			α2(VI)			α1(VI)		
	N1	C1	C2	N1	C1	C2	N1	C1	C2
α3(VI)N2	21(2.8)	20(5.7)	21(3.0)	23(4.4)	27(5.1)	21(6.0)	22 (3.9)	28 (6.5)	21 (4.4)
α3(VI)N1	–	21(5.8)	21(2.9)	18(2.2)	21(4.4)	19(4.6)	20 (0.9)	19 (2.3)	23 (4.8)
α3(VI)C1	–	–	22(2.1)	21(1.9)	25(5.7)	21(5.6)	21 (1.3)	21 (5.9)	19 (4.2)
α3(VI)C2	–	–	–	19(1.3)	21(8.1)	19(5.0)	17 (1.8)	25 (6.8)	19 (7.3)
α2(VI)N1	–	–	–	–	26(4.8)	26(4.3)	35(14.2)	24 (5.8)	18 (3.5)
α2(VI)C1	–	–	–	–	–	21(6.4)	26 (5.4)	42(19.3)	24 (6.2)
α2(VI)C2	–	–	–	–	–	–	19 (3.2)	17 (5.5)	22(10.6)
α1(VI)N1	–	–	–	–	–	–	–	29 (8.6)	24 (6.5)
α1(VI)C1	–	–	–	–	–	–	–	–	23 (6.7)

Determined with the program PIRALIGN and expressed as % identical residues and alignment scores given in SD units. Values of 3.09, 4.75, 6 and 9.26 SD correspond to probabilities of 10^{-3}, 10^{-6}, 10^{-9} and 10^{-20} respectively, for being similar by chance. Sequence data were taken from Figure 2 and from Chu *et al.* (1989) for the human α1(VI) and α2(VI) chains.

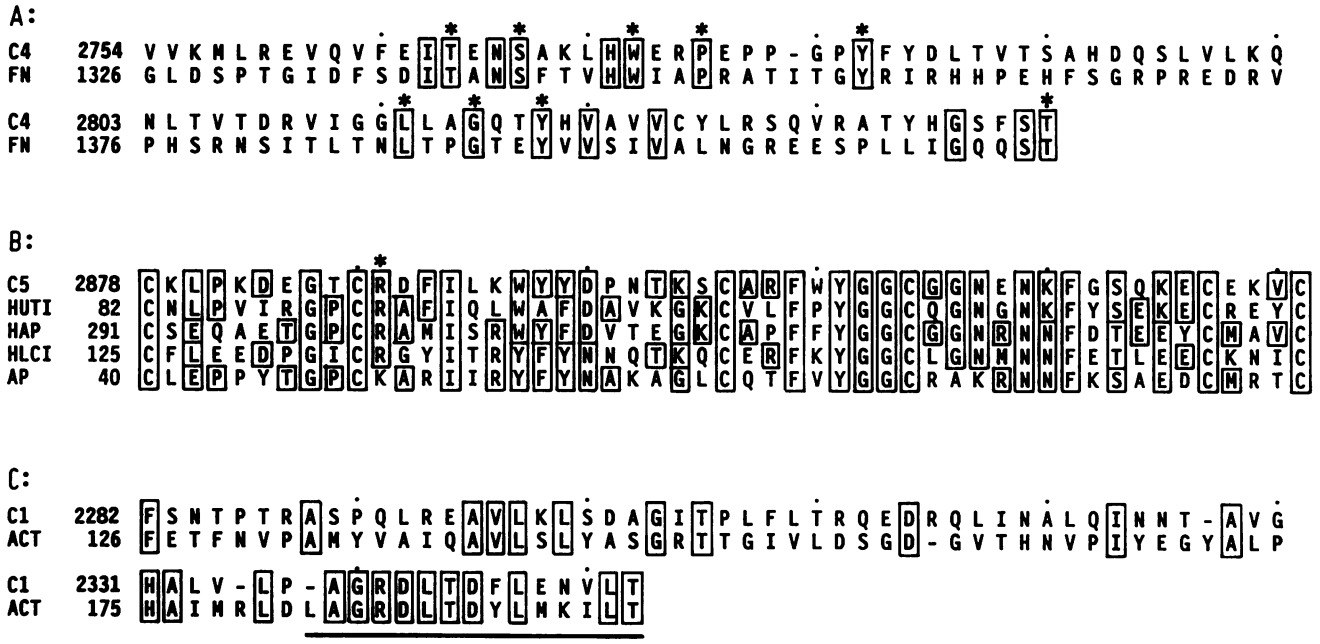

Fig. 5. Comparison of the C-terminal globular domains C1, C4 and C5 of α3(VI) chain with homologous protein sequences. (**A**) FN denotes the ninth type III repeat in human fibronectin (Kornblihtt *et al.*, 1985). Asterisks above the sequence denote residues which are almost invariant in the type III repeats found in human fibronectin and chick tenascin/cytotactin (Kornblihtt *et al.*, 1985; Jones *et al.*, 1989). (**B**) Comparison with bovine aprotinin (AP) (Kassell and Laskowski, 1965), the human urinary trypsin inhibitor (HUTI) (Wachter and Hochstrasser, 1981) and similar domains found in a human amyloid protein (HAP) (Ponte *et al.*, 1988) and a lipoprotein associated coagulation inhibitor (HLCI) (Wun *et al.*, 1988). An asterisk denotes R or K within the active site. (**C**) Comparison of the C-terminal end of C1 with an internal sequence of human α-actin (ACT) (Hanauer *et al.*, 1983). A 14 residue sequence motif found in several actin binding proteins (Tellam *et al.*, 1989) is underlined. Identical residues are boxed; numbers refer to the positions of Figure 2 and those in the compared protein sequences.

segment of α-actin (Hanauer *et al.*, 1983) and a 78-residue sequence located at the C-terminal end of the α3(VI) chain C1 domain (Figure 5C).

Three other domains C3, C4 and C5 are predicted for the most distal sequence of the α3(VI) C-terminal globule. They lack any relationship to the other globular domains of the α3(VI) chain and are not found in the α1(VI) and α2(VI) chains. The 122-residue domain C3 shows a unique composition with Lys, Pro, Thr, Ala and Val contributing to 90% of the sequence. This domain is characterized by rather invariant 22-fold repeats of the Lys-Pro sequence (KP repeat) which are separated from each other by two to five other residues (Figure 6). A similar (37% identity), but somewhat larger domain with less variable repeats (spacings mainly Thr-Thr-Thr), is found in the salivary glue protein sgs-3 of *Drosophila* (Garfinkel *et al.*, 1983). Further significant identities (22–31%) were found for several human proline rich salivary proteins (Ayen *et al.*, 1984; Kauffman *et al.*, 1986), calpain inhibitor (Emori *et al.*, 1987) and several H1 histones (Golas and Wells, 1985; Ohe *et al.*, 1986). Such identities may, however, reflect the uniform sequence rather than structural or functional relationships.

A 89 residue segment C4 directly adjacent to C3 has features typical of type III domains originally detected in fibronectin (Peterson *et al.*, 1983). This identification is mainly based on the preservation of size and several invariant

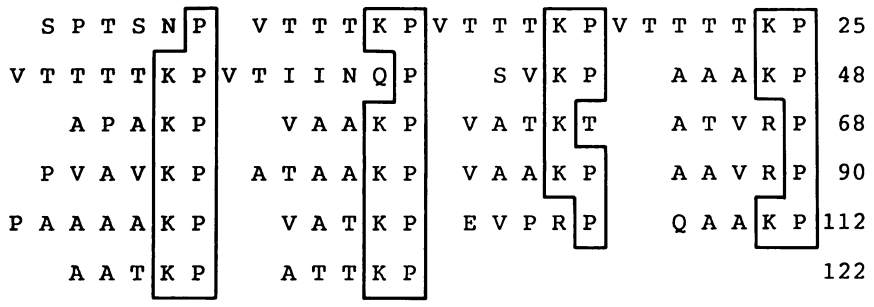

Fig. 6. Internal alignment of the repetitive sequence present in domain C3 of the α3(VI) chain. The alignment in segments of 20–25 residues followed a similar one for human proline rich salivary protein (Ayen *et al.*, 1984) keeping the most conserved KP repeat (shown in boxes) in invariant positions. The numbering is different to that of Figure 2.

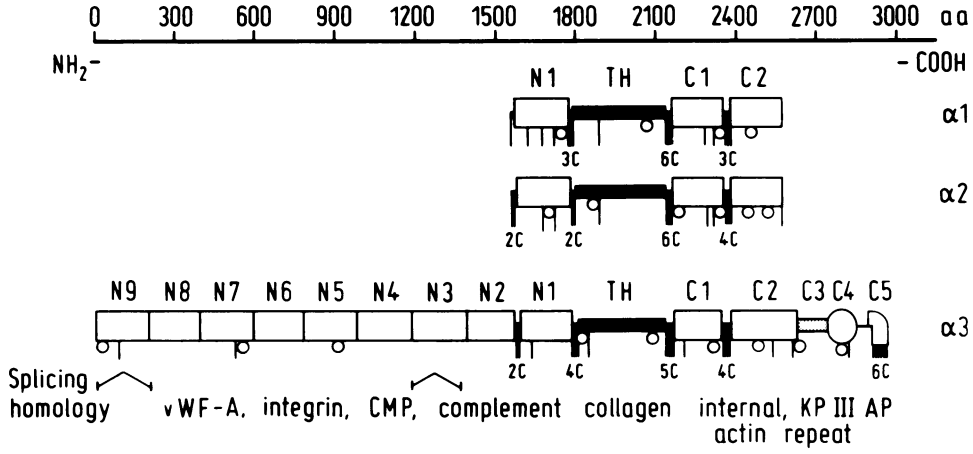

Fig. 7. Schematic diagram of the mosaic structure of the human α3(VI) chain which consists of N-terminal globular domains N9 to N1, a triple helical segment (TH) and C-terminal globular domains C1 to C5 and comparison to the domain structures of the α1(VI) and α2(VI) chains. Short horizontal lines indicate small connecting segments. Localization of N-glycosylation sites (○) and cysteine residues (vertical lines) with additional numbers when clustered are indicated. Most obvious homologies to other proteins are indicated, including A domains of von Willebrand Factor (vWF-A), cartilage matrix protein (CMP), type III domain of fibronectin (III) and aprotinin (AP). KP repeat refers to a repetitive sequence similar to those of salivary proteins. The domain organization of human α1(VI) and α2(VI) chains was reported by Chu *et al.* (1989). The two splice sites refer to those shown for several clones in Figure 1. The scale of amino acid numbers (aa) is shown above.

amino acid residues (Figure 5A) since the identity (18%) to the best fitting domain of human fibronectin (Kornblihtt *et al.*, 1985) is rather low. The identification of several similar type III domains in the extracellular matrix protein tenascin/cytotactin has been based on analogous considerations (Jones *et al.*, 1989; Spring *et al.*, 1989).

The last 70 residue segment of the C-terminal globule (domain C5) shows six cysteines located within 51 residues and 42% identity (Figure 5B) to a comparable segment of the trypsin inhibitor aprotinin (Kassell and Laskowski, 1965). Even higher identities up to 50% were found for some related proteins, e.g. human urinary trypsin inhibitor (Wachter and Hochstrasser, 1981) and a lipoprotein associated coagulation inhibitor (Wun *et al.*, 1988) among many other serine protease inhibitors (Laskowski and Kato, 1980). Homologous proteins include also a recently described amyloid protein (Ponte *et al.*, 1988; Tanzi *et al.*, 1988; Kitaguchi *et al.*, 1988). The identity to α3(VI) domain C5 is not restricted to just the Cys residues but involves several more invariant Gly, Asn, Tyr, Phe residues (Figure 5B) and a basic amino acid in the active inhibitor (P1) site (position 11 of C5 in Figure 5B, corresponding to position 15 in aprotinin).

Discussion

The whole amino acid sequence of the mature human collagen α3(VI) chain (2943 positions) predicts a molecular mass of ~340 kd (including some glycosylation). This is a 30% higher mass than estimated previously from electrophoretic analyses (Engvall *et al.*, 1986; Trüeb and Winterhalter, 1986; Colombatti *et al.*, 1987). It could indicate a low accuracy of the electrophoretic assay as found before in similar molecular mass estimates for fibronectin (Kornblihtt *et al.*, 1985; Ruoslahti, 1988) or limited processing of the α3(VI) chain. Results of Edman degradation (Figure 2) demonstrated the presence of globular domains N1 to N8 and C1 (Figure 7) in the tissue form of type VI collagen. The N-terminal domains very likely correspond to a 160 kd fragment released from the α3(VI) chain by collagenase (Trüeb and Winterhalter, 1986), while the same study failed to identify a 90–100 kd fragment which could correspond to the whole C-terminal globule. These findings suggest processing of the C-terminal globule including the release of the putative inhibitor domain C5 despite the failure to show substantial processing in cell culture (Colombatti and Bonaldo, 1987). No such processing

has been found, however, for the α1(VI) and α2(VI) chain (Chu *et al.*, 1989).

Further differences between the three chains of type VI collagen exist in their sizes (Figure 7). They share a triple helical segment of identical length and three similar globular domains, N1, C1 and C2. The similarity of these domains is larger between α1(VI) and α2(VI) chains when compared with the α3(VI) chain (Table I). Assuming that all three chains originated from a common ancestral precursor, this observation suggests an early divergence of the α3(VI) from the α1(VI) and α2(VI) chains during evolution. This could be reflected by a different chromosomal localization in the human genome being 2q37 for the α3(VI) and 21q223 for both α1(VI) and α2(VI) chains (Weil *et al.*, 1988). Yet, the other eight domains, N2 to N9, show a higher internal homology compared with N1, C1 and C2 and may have evolved from a more recent multiplication event.

These various globular domains shared by all three chains show distinct homology to the A domains of von Willebrand Factor for which binding to collagen, heparin and cell surface structures has been reported (Titani and Walsh, 1988). This has raised the possibility that the similar type VI collagen domains have a functional role in the assembly of microfibrils and their heterotypic binding to large collagen fibrils and basement membrane collagen (Bonaldo *et al.*, 1989; Chu *et al.*, 1989; Koller *et al.*, 1989). Their abundance in the α3(VI) chain, however, also indicates that not all of them may have binding activity and alternatively might serve as spacer elements.

Three more domains unique to the α3(VI) chain, C3 to C5 (Figure 7), complete the mosaic structure of type VI collagen and could be important for different functions. The KP repeat C3 may, as in other salivary proteins (Beckendorf and Kafatos, 1976; Korge, 1977; Garfinkel *et al.*, 1983), be utilized for *O*-glycosylation. Another 31 residue segment rich in Pro, Thr and Ser (see Figure 2, positions 2842–2873) preceeding C5 could also be used for the same purpose. The type III repeat of C4 is found 15–17 times in fibronectin (Kornblihtt *et al.*, 1985; Ruoslahti *et al.*, 1988), where they form rod-like elements with binding sites for heparin and cells. The characteristic cell binding sequence RGD is, however, not found in C4; rather it is present in the triple helical segment of all three type VI collagen chains (Figure 2) (Chu *et al.*, 1988; Aumailley *et al.*, 1989).

The best functional prediction, based on sequence comparisons, exists for the inhibitor-related domain C5. The presence of an Arg in its presumed active site indicates that the domain, if active, should bind to proteases with trypsin like specificity. A similar inhibitor motif present in a predicted amyloid precursor was shown recently by recombinant technology to be active against trypsin (Kitaguchi *et al.*, 1988). Most inhibitors of this type are expressed as small proteins containing no extra domains (Laskowski and Kato, 1980), underscoring the possibility that C5 could be released from type VI collagen in tissues. These possibilities can be studied now by using recombinant products and antibodies raised against them.

A further variability of human the α3(VI) chain structure was suggested by Northern hybridization which showed two to four bands in the size range 8.3–9.2 kb (Chu *et al.*, 1987). The upper value could also be an underestimate because of the predicted maximal size (9.9 kb) shown here. The variations could be due to the use of different polyadenylation signals, yet three independent clones (FM12, FM20 and FM22) showed the same 3′ non-coding region of 759 nt, with only one typical polyadenylation site (Figure 2). Other, more 5′ clones show one or two deletions not seen in other clones (Figure 1). These deletions eliminate 167 amino acid residues of domain N3 and almost entirely domain N9. This suggests alternative splicing of the α3(VI) mRNA. Structural variants which differ in the C-terminus were also predicted for the α2(VI) chain (Chu *et al.*, 1989). They were shown to arise by the mutually exclusive use of two exons encoding domain C2 and the 3′ non-coding region (Chu *et al.*, 1990). No evidence exists so far for similar variability at the 3′ end of the α1(VI) and α3(VI) chain mRNA. The functional consequences and tissue specificity of such variations remain to be determined.

Materials and methods

Isolation of cDNA clones and nucleotide sequencing

Total RNA was isolated from human fibroblasts (GM 3349) (Coriell Institute for Medical Research, Camden, NJ) by acid guanidine thiocyanate–phenol–chloroform extraction. Poly(A)$^+$ was then selected on an oligo(dT)–cellulose column (Aviv and Leder, 1972). Three λZAP cDNA libraries were prepared from 5 μg each of poly(A)$^+$ RNA using either oligo(dT)12–18 or mixed hexanucleotides (Pharmacia, Piscataway, NJ) as primers by established procedures (Gübler and Hoffmann, 1983) provided in cDNA synthesis kits (Bethesda Research Laboratories, Gaithersburg, MD and Pharmacia, Piscataway, NJ). The double stranded cDNA was size selected by chromatography on a Bio-Gel A-50m column, ligated with *Eco*RI linker or adaptor, and cloned into the *Eco*RI site of λZAPII vector (Stratagene, La Jolla, CA). An additional primer extension library was constructed using as primer an oligonucleotide (19 mer) synthesized according to the cDNA sequence (SP308, Figure 1). The specific primer was annealed to the poly(A)$^+$ RNA at 65°C for 30 min, prior to the synthesis of the first strand. The unamplified cDNA libraries were screened with ^{32}P-labeled fragments from previously isolated cDNA clones by standard procedures (Rigby *et al.*, 1977; Benton and Davis, 1977). Following plaque purification, the positive cDNA clones in Bluescript plasmids were excised from the phage vectors by infecting with single strand helper phage R408, according to the protocol provided by Stratagene.

The cDNA clones were sequenced by dideoxynucleotide chain termination method (Sanger *et al.*, 1977) using [^{35}S]thio-dATP (New England Nuclear, Boston, MA) and the modified T7 polymerase (Sequenase kit, USB, Cleveland, OH). Some sequences were determined with fluorescent M13 primers and analyzed using an automated DNA sequencer (Applied Biosystems, Foster City, CA). The double stranded plasmids were first sequenced from the 5′ and 3′ directions using two primers flanking the Bluescript vector (Chen and Seeburg, 1985). To obtain the complete sequence, the cDNA inserts were digested with appropriate restriction enzymes and then subcloned into the single stranded M13 vectors. The single stranded DNA templates were sequenced by using M13 universal primers (Amersham, Arlington Heights, IL) or specific primers complementary to internal cDNA sequences, which were synthesized with a DNA synthesizer (Dupont, Coder 300).

Peptide isolation and Edman degradation

A 30 kd fragment NC1 was isolated from a collagenase digest of pepsin solubilized human type VI collagen (Odermatt *et al.*, 1983) and recently shown to comprise a portion of the C-terminal globule of α3(VI) chain (Kuo *et al.*, 1989). The reduced and alkylated fragment was cleaved with TPCK–trypsin (2 h, 37°C) and the digest separated by HPLC on a reversed phase column (Chu *et al.*, 1987). Another set of small peptides were obtained from a Lys-C protease digest of collagenase resistant fragments obtained from a 140 kd pool of human type VI collagen chains (Chu *et al.*, 1989). The various purified peptides were subjected to Edman degradation (15–30 cycles) in a gas phase sequencer (model 470A, Applied Biosystems) as previously described (Chu *et al.*, 1987, 1989).

Comparison of amino acid sequences

The deduced amino acid sequence was searched in the protein sequence data bank (MIPSX Data base, F.Pfeiffer, Martinsried Institute of Protein Sequences) for homologous proteins by the FASTP program (Lipman and Pearson, 1985). Alignment of two particular sequences was achieved with

the PIRALIGN program (Dayhoff *et al.*, 1983; George *et al.*, 1986) using the mutation data matrix with a bias of +6 and a gap penalty of 6. The alignments were compared with 100 random permutations to obtain the alignment scores. Multiple sequences were then aligned based on these comparisons.

Acknowledgements

We thank Loretta Renkart, Cynthia Bohan, Bruce Donaldson and Kerry Maddox for excellent technical assistance. The study was supported by the National Institutes of Health (grants AR 38912, AR 38923, AR 38188 and AR 39740), the Deutsche Forschungsgemeinschaft (SFB 266) and Shriners Hospital for Crippled Children.

References

Aumailley,M., Mann,K., von der Mark,H. and Timpl,R. (1989) *Exp. Cell Res.*, **181**, 463–474.
Aviv,H. and Leder,P. (1972) *Proc. Natl. Acad. Sci. USA*, **69**, 1408–1412.
Ayen,E., Lyons,K.M., McGonigal,T., Barett,N.L., Clements,L.S., Maeda,N., Vanin,E.F., Carlson,M.D. and Smithies,O. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 5561–5565.
Beckendorf,S.K. and Kafatos,F.C. (1976) *Cell*, **9**, 365–373.
Benton,W.D. and Davis,R.W. (1977) *Science*, **196**, 180–182.
Bonaldo,P., Russo,V., Bucciotti,F., Bressan,G.M. and Colombatti,A. (1989) *J. Biol. Chem.*, **264**, 5575–5580.
Caput,D., Beutler,B., Hartog,K., Thayer,R., Brown-Shimer,S. and Cerami,A. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 1670–1674.
Chen,E.Y. and Seeburg,P.H. (1985) *DNA*, **4**, 165–170.
Chu,M.-L., Mann,K., Deutzmann,R., Pribula-Conway,D., Hsu-Chen,C.-C., Bernard,M.P. and Timpl,R. (1987) *Eur. J. Biochem.*, **168**, 309–317.
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., Pan,T.-C., Conway,D., Saitta,B., Stokes,D., Kuo,H.-J., Glanville,R.W., Timpl,R., Mann,K. and Deutzmann,R. (1990) *Ann. N.Y. Acad. Sci.*, in press.
Colombatti,A. and Bonaldo,P. (1987) *J. Biol. Chem.*, **262**, 14461–14466.
Colombatti,A., Bonaldo,P., Ainger,K., Bressan,G.M. and Volpin,D. (1987) *J. Biol. Chem.*, **262**, 14454–14460.
Dayhoff,M.D., Barker,W.C. and Hunt,L.T. (1983) *Methods Enzymol.*, **91**, 524–545.
Emori,Y., Kawasaki,H., Imajoh,S., Imahori,K. and Suzuki,K. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 3590–3594.
Engvall,E., Hessle,H. and Klier,G. (1986) *J. Cell Biol.*, **102**, 703–710.
Furthmayr,H., Wiedemann,H., Timpl,R., Odermatt,E. and Engel,J. (1983) *Biochem. J.*, **211**, 303–311.
Garfinkel,M.D., Pruitt,R.E. and Meyerowitz,E.M. (1983) *J. Mol. Biol.*, **168**, 765–789.
George,D.G., Barker,N.C. and Hunt,L.T. (1986) *Nucleic Acids Res.*, **14**, 11–15.
Golas,L.S. and Wells,J.R.E. (1985) *Nucleic Acids Res.*, **13**, 585–594.
Gübler,U. and Hoffmann,B.J. (1983) *Gene*, **25**, 263–269.
Hanauer,A., Levin,M., Heilig,R., Daegelen,D., Kahn,A. and Mandel,J.L. (1983) *Nucleic Acids Res.*, **11**, 3503–3516.
Heckmann,M., Aumailley,M., Hatamochi,A., Chu,M.-L., Timpl,R. and Krieg,T. (1989) *Eur. J. Biochem.*, **182**, 716–726.
Jander,R., Troyer,D. and Rauterberg,J. (1984) *Biochemistry*, **23**, 3675–3681.
Jones,F.S., Hoffman,S., Cunningham,B.A. and Edelman,G.M. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 1905–1909.
Kassell,B. and Laskowski,M. (1965) *Biochem. Biophys. Res. Commun.*, **20**, 463–468.
Kauffman,D., Hofmann,T., Bennick,A. and Keller,P. (1986) *Biochemistry*, **25**, 2387–2392.
Kiss,I., Deak,F., Holloway,R.G., Delius,H., Mebust,K.A., Frimberger,E., Argraves,W.S., Tsonis,P.A., Winterbottom,N. and Goetinck,P.F. (1989) *J. Biol. Chem.*, **264**, 8126–8134.
Kitaguchi,N., Takahashi,Y., Tokushima,Y., Shiojiri,S. and Ito,H. (1988) *Nature*, **331**, 530–532.
Koller,E., Winterhalter,K.H. and Trüeb,B. (1989) *EMBO J.*, **8**, 1073–1077.
Korge,G. (1977) *Dev. Biol.*, **58**, 339–355.
Kornblihtt,A.R., Umeyawa,K., Vibe-Pedersen,K. and Baralle,F.E. (1985) *EMBO J.*, **4**, 1755–1759.
Kuo,H.-J., Keene,D.R. and Glanville,R.W. (1989) *Biochemistry*, **28**, 3757–3762.
Laskowski,M. and Kato,I. (1980) *Annu. Rev. Biochem.*, **49**, 593–626.
Lipman,D.J. and Pearson,W.R. (1985) *Science*, **227**, 1435–1441.
Odermatt,E., Risteli,J., van Delden,V. and Timpl,R. (1983) *Biochem. J.*, **211**, 295–302.
Ohe,Y., Hayashi,H. and Iwai,K. (1986) *J. Biochem.*, **100**, 359–368.
Peterson,T.E., Thogersen,H.C., Skorstengaard,K., Vibe-Pedersen,K., Sahl,P., Sottrup-Jensen,L. and Magnusson,S. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 137–141.
Ponte,P., Gonzalez-DeWhitt,P., Schilling,J., Miller,J., Hsu,D., Greenberg,B., Davis,K., Wallace,W., Lieberburg,I., Fuller,F. and Cordell,B. (1988) *Nature*, **331**, 525–527.
Rigby,P.W.J., Dieckmann,M., Rhodes,C. and Berg,P. (1977) *J. Mol. Biol.*, **113**, 237–251.
Ruoslahti,E. (1988) *Annu. Rev. Biochem.*, **57**, 375–413.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
Shaw,G. and Kamen,R. (1986) *Cell*, **46**, 659–667.
Spring,J., Beck,K. and Chiquet-Ehrisman,R. (1989) *Cell*, **59**, 325–334.
Tanzi,R.E., McClatchey,A.I., Lamperti,E.D., Villa-Komaroff,L., Gusella,J.F. and Neve,R.L. (1988) *Nature*, **331**, 528–530.
Tellam,R.L., Morton,D.J. and Clarke,F.M. (1989) *Trends Biochem. Sci.*, **14**, 130–133.
Timpl,R. and Engel,J. (1987) In Mayne,R. and Burgeson,R.E. (eds), *Structure and Function of Collagen Types*. Academic Press, Orlando, FL, pp. 105–143.
Titani,K. and Walsh,K.A. (1988) *Trends Biochem. Sci.*, **13**, 94–97.
Titani,K., Kumar,S., Takio,K., Ericsson,L.H., Wade,R.D., Ashida,K., Walsh,K.A., Chopek,M.W., Sadler,J.E. and Fujikawa,K. (1986) *Biochemistry*, **25**, 3171–3184.
Trüeb,B. and Winterhalter,K. (1986) *EMBO J.*, **5**, 2815–2819.
von der Mark,H., Aumailley,M., Wick,G., Fleischmajer,R. and Timpl,R. (1984) *Eur. J. Biochem.*, **142**, 493–502.
von Heijne,G. (1985) *J. Mol. Biol.*, **184**, 99–105.
von Heijne,G. (1986) *Nucleic Acids Res.*, **14**, 4683–4690.
Wachter,E. and Hochstrasser,K. (1981) *Hoppe-Seyler's Z. Physiol. Chem.*, **362**, 1351–1355.
Weil,D., Mattei,M.-G., Passage,E., Van Cong,N., Pribula-Conway,D., Mann,K., Deutzmann,R., Timpl,R. and Chu,M.-L. (1988) *Am. J. Hum. Genet.*, **42**, 435–445.
Wun,T.-Z., Kretzmer,K.K., Girard,I.J., Miletich,J.P. and Broze,G.J. (1988) *J. Biol. Chem.*, **263**, 6001–6004.

Received on October 9, 1989; revised on November 17, 1989