The EMBO Journal vol.8 no.4 pp.1073 - 1077, 1989

The globular domains of type VI collagen are related to the collagen-binding domains of cartilage matrix protein and von Willebrand factor

Erich Koller, Kaspar H.Winterhalter and Beat Trueb

Laboratorium für Biochemie I, ETH Zentrum, CH-8092 Zürich, Switzerland

Communicated by R.Friis

Type VI collagen is a transformation-sensitive glycoprotein of the extracellular matrix of fibroblasts. We have isolated and sequenced several overlapping cDNA clones (4153 bp) which encode the entire α2 subunit of chicken type VI collagen. The deduced amino acid sequence predicts that the α2(VI) polypeptide consists of 1015 amino acid residues that are arranged in four domains: a hydrophobic signal peptide of 20 residues, an amino-terminal globular domain of 228 residues, a collagenous segment of 335 residues and a carboxy-terminal globular domain of 432 residues. The collagenous domain contains seven Arg-Gly-Asp tripeptide units, some of which are likely to be used as cell-binding sites. The globular domains contain three homologous repeats with an average length of 180 amino acid residues. These repeats show a striking similarity to the collagen-binding motifs found in von Willebrand factor and cartilage matrix protein. We therefore speculate that the globular domains of the α2(VI) polypeptide may interact with collagenous structures.

Key words: collagen type VI/collagen-binding domain/cartilage matrix protein/von Willebrand factor/Mac-1

Introduction

Type VI collagen is a major constituent of the extracellular matrix of most tissues (see Timpl and Engel, 1987, for a review). The protein is composed of three different polypeptide chains that form a heterotrimeric molecule with a short triple helix and large globular domains at both ends. Type VI collagen is synthesized and secreted by mesenchymal cells. In the extracellular matrix it forms small microfilaments which are often associated with the large banded fibres of the interstitial collagens (Bruns *et al.*, 1986; Sakai *et al.*, 1986).

Our interest has focused on type VI collagen because this protein is specifically down-regulated in fibroblasts transformed by DNA or RNA tumour viruses or in cells derived from spontaneous tumours (Carter, 1982a; Schreier *et al.*, 1988). A reduction in the synthesis of type VI collagen might have dramatic effects on the adhesive properties of transformed cells as the protein is known to promote cell attachment in a way similar to fibronectin (Carter, 1982b). Inhibition of type VI collagen synthesis appears to occur at the transcriptional level since transformed cells do not contain any type VI collagen specific mRNA (Trueb *et al.*, 1985, 1989). When fibroblasts are infected with a strain of Rous sarcoma virus that contains a temperature-sensitive mutation in the transforming gene *src*, the cells produce type VI collagen at the restrictive, but not at the permissive temperature (Schreier *et al.*, 1988). If such cells are shifted from the permissive to the restrictive temperature, the synthesis of the individual subunits of type VI collagen is co-ordinately induced. The activity of a single oncogene is therefore sufficient to inhibit type VI collagen expression. To study the expression at the molecular level we have begun with the cloning of specific cDNA probes for type VI collagen. In a preliminary publication we reported the sequence of a cDNA clone corresponding to the central part of the α2(VI) polypeptide chain (Trueb *et al.*, 1989). The collagenous domain of this chain was shown to consist of 335 amino acid residues and to contain only one interruption in the repetitive Gly-X-Y sequence.

Here we report the isolation of several overlapping cDNA clones which together encode the entire α2 subunit of chicken type VI collagen. The nucleotide sequence predicts a polypeptide of 1015 amino acid residues with two globular and one collagenous domain. Of particular interest are three homologous repeats in the globular domains that show a striking similarity to the collagen-binding motifs found in von Willebrand factor and cartilage matrix protein. This allows speculation that the globular domains of type VI collagen are able to interact with collagenous structures.

Results

Isolation of cDNA clones

Screening of a chicken cDNA library with a cDNA probe corresponding to the central part of the α2(VI) polypeptide chain (Trueb *et al.*, 1989) led to the isolation of eight overlapping cDNA clones (Figure 1). Only one of these clones (5a) turned out to be identical with the probe used for screening. The sequences of the other cDNA clones were established by the dideoxy technique. Altogether the clones span 4153 bp and encode the entire α2 subunit of chicken type VI collagen (Figure 2).

Analysis of the nucleotide sequence

The first and second ATG codon at the 5′ end of the nucleotide sequence are followed by in-frame stop codons. The nucleotides surrounding the third ATG codon do not conform to the consensus sequence for initiation of translation in vertebrates (Kozak, 1987). Only the fourth ATG codon occurs in a nucleotide surrounding that is consistent with a typical start site for translation. This initiation codon is followed by an open reading frame (ORF) of 3045 bp. The beginning and end of this reading frame are also indicated by analysis of the nucleotide sequence with the algorithm TESTCODE which is based on measurements of the compositional constraint of a sequence at every third position (Fickett, 1982). Following the ORF there is a 3′

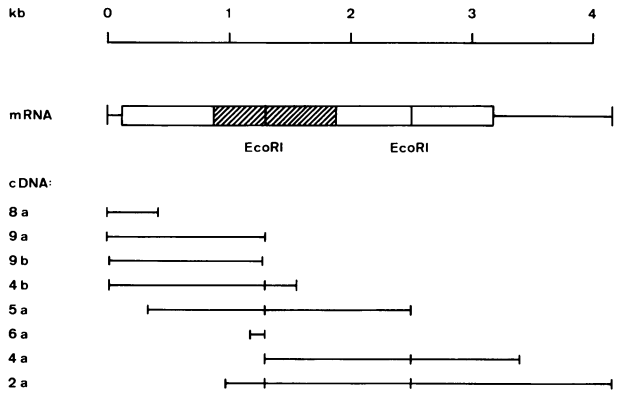

Fig. 1. Arrangement of the cDNA clones relative to the mRNA of α2(VI). The ORF is boxed, the collagenous region is hatched, the internal *Eco*RI restriction sites are indicated.

untranslated region of 973 bp. Two polyadenylation signals are found in this region 32 and 24 bp upstream from the site of poly(A) addition.

Analysis of the amino acid sequence

The amino acid sequence derived from the nucleotide sequence predicts that the α2(VI) polypeptide consists of 1015 amino acid residues. The sequence starts with a typical signal peptide of 20 residues. This signal peptide represents the most hydrophobic part of the entire protein on a hydrophilicity plot according to Kyte and Doolittle (1982). The rules of von Heijne (1986) predict that the peptide is cleaved between alanine (position −1) and glutamine (position 1) with a score of 6.9. The mature polypeptide starts therefore with a glutamine residue which can spontaneously convert into pyroglutamic acid and block the N terminus of the polypeptide chain as observed with many other extra-

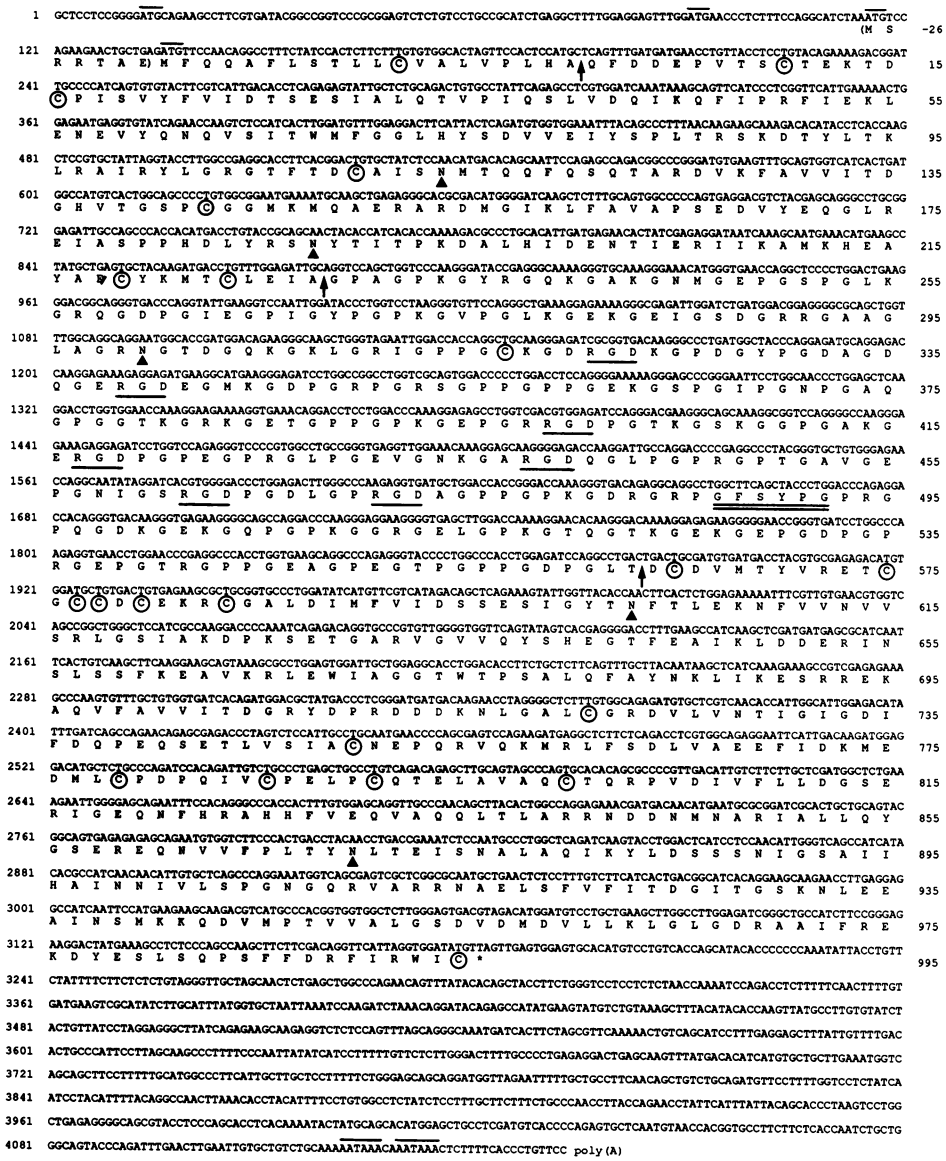

Fig. 2. Nucleotide sequence of the α2(VI) cDNA and derived amino acid sequence. Possible initiation codons ATG and the polyadenylation signals are indicated by straight lines. The extension of the signal peptide arising by initiation at an upstream initiation codon is given in parentheses. Arrows indicate the cleavage site of the signal peptide and the beginning and end of the collagenous region. Putative cell-binding sites RGD are underlined. The imperfection in the repetitive G-X-Y sequence is given by a double line. Cysteine residues are encircled, possible glycosylation sites NXT/S are indicated by triangles.

cellular matrix proteins. The mature polypeptide (995 amino acid residues, M_r 106 141) is composed of three different domains: an amino-terminal globular domain of 228 residues, a collagenous segment of 335 residues and a carboxy-terminal globular domain of 432 residues. The amino-terminal globular domain (M_r 25 887) contains two sites for the attachment of N-linked carbohydrates and six cysteine residues. The collagenous segment (M_r 31 956) shows one interruption in the repetitive Gly-X-Y sequence and contains one cysteine residue and one glycosylation signal. Of particular interest are seven Arg-Gly-Asp tripeptide units in this domain that may function as cell adhesion sites. The carboxy-terminal domain (M_r 48 352) contains two glycosylation signals and 13 cysteine residues. Six of these cysteine residues occur in a cluster following the collagenous region. It is likely that some of these residues connect the three subunits of a type VI collagen molecule by the formation of disulphide bonds.

Internal homologies

The amino acid sequence of the α2(VI) chain was examined for internal homologies by dot matrix analysis (Figure 3). Not surprisingly, numerous stretches of homology occur within the collagenous region (residues 229–563). In addition there are three homologous repeats with an average length of 180 amino acid residues, one in the amino-terminal globular domain (D1 residues 1–176) and two in the carboxy-terminal globular domain (D2 residues 569–745 and D3 790–964). The overall degree of identity between the three homologous repeats is 19–23%; if conservative amino acid replacements are included, this number increases to 43–48% (Figure 4). Evaluation with the program ALIGN shows that the homology between the three domains is of high statistical significance (alignment score in SD: D1/D2, 5.6; D1/D3, 5.6; D2/D3, 7.9). Thus, the probability that the alignments have occurred by chance is $< 10^{-7}$. Each repeat unit is composed of two smaller subdomains of high homology spaced by ~90 amino acid residues of lesser homology. The degree of identity between the smaller subdomains reaches 45% (Figure 5). A computer search of published amino acid sequences reveals similar sequence motifs of 180–200 amino acid residues in von Willebrand factor (domains A1, A2 and A3), in cartilage matrix protein (CMP-1, CMP-2) and in the leukocyte adhesion receptors Mac-1 and p150,95. All these domains share a significant degree of homology (~20%). In each case the homology is most pronounced in the region of two smaller subdomains which are spaced by ~80 amino acid residues (Figure 5). In the case of type VI collagen the homologous repeats are linked to each other and to the collagenous domain by short cysteine-rich segments. Together, the homologous repeats, the collagenous domain and the cysteine-rich segments build up the entire α2(VI) polypeptide chain (Figure 6).

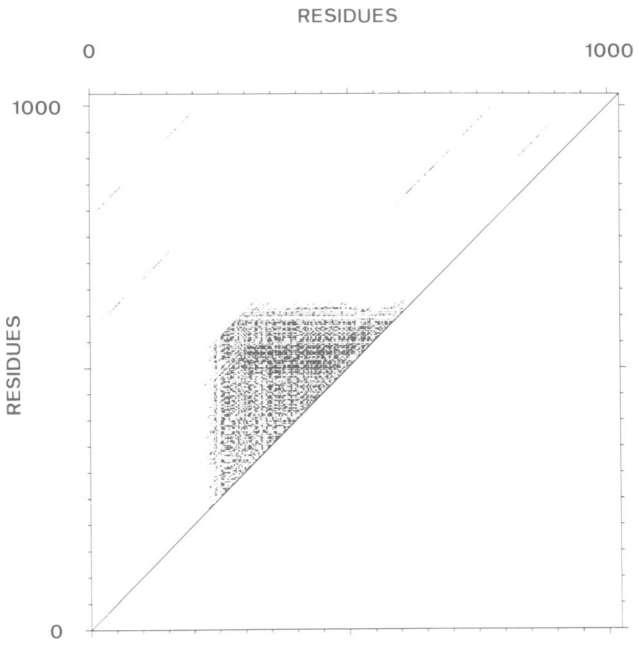

Fig. 3. Dot matrix analysis of the α2(VI) amino acid sequence. The residues of the signal peptide are included in this comparison.

Discussion

We have determined the complete primary structure of the α2 subunit of chicken type VI collagen by analysis of several overlapping cDNA clones. The cDNA clones contain a short poly(A) tail and cover a total of 4153 bp. This number corresponds well with the size of the mRNA for the α2(VI)

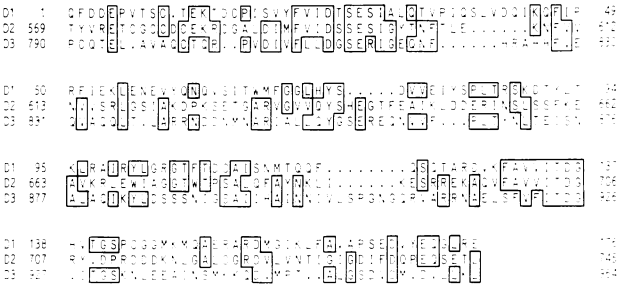

Fig. 4. Alignment of the three homologous repeats of the α2(IV) polypeptide. Identical residues are boxed.

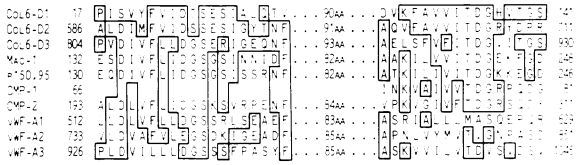

Fig. 5. Comparison of the homologous repeats of type VI collagen (Co16 D1, D2, D3) with the collagen-binding domains of the leukocyte adhesion receptors Mac-1 and p150,95, the cartilage matrix protein (CMP-1, CMP-2) and von Willebrand factor (vWF A1, A2, A3). Common residues between the sequence of α2(VI) and the rest of the proteins are boxed. For a more detailed compilation of the sequence data of Mac-1, p150,95, CMP and vWF cf. Pytela (1988) and Corbi *et al.* (1988).

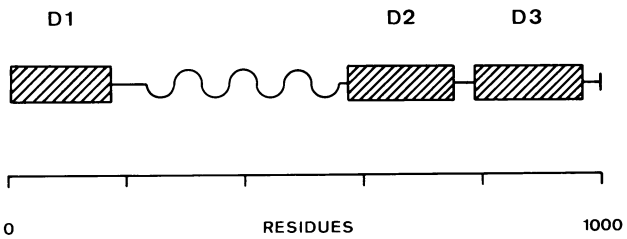

Fig. 6. Domain structure of the α2(VI) polypeptide. The putative collagen-binding domains are given as hatched boxes, the collagenous region is indicated by a wavy line.

polypeptide (4200 nucleotides) that has been determined by Northern blot hybridization (Trueb *et al.*, 1989).

The nucleotide sequence predicts that the α2(VI) polypeptide is composed of 1015 amino acid residues. At the amino terminus the polypeptide contains a typical signal peptide of 20 residues. After removal of this signal peptide the mature α2(VI) chain consists of 995 amino acid residues which are arranged in three distinct domains: an amino-terminal globular domain of 228 residues, a collagenous segment of 335 residues and a carboxy-terminal globular domain of 432 residues. Under the assumption that all potential glycosylation sites Asn-X-Thr/Ser are occupied with oligosaccharides (average M_r 2500) the domains would have a molecular mass of 30 900, 34 500 and 53 400 respectively. This is in good agreement with the values of 32 000, 38 000 and 53 000 that have been determined for the fragments C3, α2(VI)-pepsin and C2 by SDS−PAGE (Gibson and Cleary, 1985). C2 and C3 represent two of the three globular fragments obtained from type VI collagen by cleavage with bacterial collagenase, α2(VI)-pepsin is one of the collagenous fragments obtained by pepsin treatment.

Of particular interest are seven Arg-Gly-Asp tripeptide sequences found in the collagenous domain of the α2(VI) chain (cf. Trueb *et al.*, 1989). Polypeptides containing this tripeptide unit have been shown to interact with cell surface receptors for fibronectin and other extracellular matrix proteins (Pytela *et al.*, 1986). Arg-Gly-Asp sequences may therefore play an important role in cell adhesion. It is likely that some of the Arg-Gly-Asp units of the α2(VI) polypeptide are used as cell-binding sites because specific integrin receptors have recently been identified that interact with collagenous proteins including type VI collagen (Dedhar *et al.*, 1987; Wayner and Carter, 1987). Type VI collagen might therefore represent an adhesive glycoprotein similar to fibronectin.

The globular domains of the α2(VI) polypeptide contain three homologous repeats of ~180 amino acid residues. One of these repeats is found in the amino-terminal globular domain and two in the carboxy-terminal globular domain. The three repeats show a striking similarity to special sequence motifs found in von Willebrand factor (domains A1, A2, A3; see Titani and Walsh, 1988, for a review), in cartilage matrix protein (Argraves *et al.*, 1987) and in the leukocyte adhesion receptors Mac-1 (Pytela, 1988) and p150,95 (Corbi *et al.*, 1988). The homologous repeats of these proteins share ~20% sequence identity. In every case the homology is most pronounced in the region of two smaller subdomains that are spaced by ~90 amino acid residues. Cysteine residues are rarely found within the homologous repeats; they occur, however, frequently in the short segments that flank the repeats and link them to other sequence motifs. Type VI collagen is unusual in that the α2(VI) chain contains three of these homologous repeats separated only by the short cysteine-rich segments and by the collagenous helix. Three similar repeats are also found in the α1(VI) chain and even more occur in the α3(VI) chain (A.Colombatti, personal communication). Given the striking homology we speculate that a single primordial sequence motif existed that was duplicated during evolution and embedded in a number of proteins.

The function of the homologous repeats in type VI collagen is not known at the moment. The function of the related sequence motifs in von Willebrand factor, however, has recently been elucidated. It has been demonstrated that the domains A1 and A2 interact specifically with fibrillar collagens (Roth *et al.*, 1986; Kalafatis *et al.*, 1987; Pareti *et al.*, 1987). This collagen-binding activity is believed to be essential for the function of the protein during haemostasis. Von Willebrand factor promotes the attachment of blood platelets to the vessel wall at the site of injury by binding to specific receptors on the platelet surface and, at the same time, to collagen fibres of the subendothelial matrix. A similar collagen-binding activity has also been found in cartilage-matrix protein (Argraves *et al.*, 1987; P.Goetinck, personal communication). Since this protein consists of two homologous repeats linked only by a short cysteine-rich segment, the collagen-binding activity must be attributed to the homologous repeats. The function of the related sequences in type VI collagen could likewise be a specific interaction with collagenous structures. Preliminary studies may support this notion. The globular domains of type VI collagen that are obtained by digestion of the molecule with bacterial collagenase are able to bind to microtitre wells coated with type VI collagen (unpublished results). The natural ligand for the homologous repeats of type VI collagen, however, might be provided by the protein itself. Type VI collagen is known to form dimeric structures by the alignment of two molecules in an anti-parallel fashion with a stagger of 30 nm (Furthmayr *et al.*, 1983). In this configuration, the globular domains of one molecule are located in close proximity to the central helix of the other molecule. It is therefore likely that the globular domains may actively bind to the proximal triple helix of the other molecule. Such an interaction between globular and helical domains would offer an explanation of how two type VI collagen molecules keep properly aligned during the formation of a dimer. Furthermore, an interaction between the globular and the helical domains would lend great mechanical stability to the extended microfilaments of type VI collagen.

An intact molecule of type VI collagen contains >10 of the homologous repeats. It is therefore conceivable that one or the other repeat may reveal a specificity for type I or type III collagen. Thus, the microfilaments of type VI collagen could also interact with the banded fibres of the interstitial collagens. A close association of type VI collagen with type I collagen fibres has in fact been described (Bruns *et al.*, 1986; Sakai *et al.*, 1986). This leads to speculation that mesenchymal cells bind via their integrin receptors to the central helix of type VI collagen and that in turn type VI collagen anchors these cells to the interstitial collagen fibres via its collagen-binding domains. In order to verify the relevance of these interactions in cell adhesion and to identify the nature of the amino acid residues involved, specific modification reagents could be utilized.

Materials and methods

Screening of the cDNA library

A cDNA probe corresponding to the central part of the α2(VI) polypeptide (Trueb *et al.*, 1989) was labeled with [α-^{32}P]dCTP by the random primed oligolabeling method (Feinberg and Vogelstein, 1983). The probe was used to screen 2 × 10^5 recombinant phages of a chicken embryo cDNA library (Clontech Laboratories, Palo Alto, CA) by the plaque hybridization technique (Benton and Davis, 1977). Eight positive clones were picked and amplified as described (Maniatis *et al.*, 1982).

DNA sequencing

The cDNA inserts were cut from the phage DNA with the restriction enzyme *Eco*RI (Boehringer) and ligated into the sequencing vector M13mp19 (Pharmacia LKB Biotechnology Inc.) according to the instructions of the manufacturer. The nucleotide sequence of the cDNA inserts was determined on both strands by the dideoxy chain termination method (Sanger *et al.*, 1977) using the enzyme Sequenase (United States Biochemical Corp.) and several synthetic oligonucleotide primers.

Sequence analysis

The sequences were analysed with the software computer package of the Genetics Computer Group, University of Wisconsin, Madison, WI (Version 5.4, July 1988) and the package of the Protein Identification Resource, National Biomedical Research Foundation, Georgetown University Medical Center, Washington DC (Version 4.4, June 1987). For the dot matrix analysis the program COMPARE was used with a window setting of 50 and a stringency of 22. Homologous sequences were aligned with the program GAP. The statistical significance of the alignment was assessed with the program ALIGN. Alignment scores were determined for 100 random permutations of the aligned sequences. The number of standard deviations between the mean of the scores for the randomized comparisons and that of the actual alignment was calculated using the mutation data matrix with a bias of 2 and a gap penalty of 5.

Acknowledgements

We thank J.Köppel for assistance with the preparation of the figures. This work was supported by the Swiss National Science Foundation (grants 3100-009383 and 3130-009065).

References

Argraves,W.S., Deak,F., Sparks,K.J., Kiss,I. and Goetinck,P.F. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 464–468.
Benton,W.D. and Davis,R.W. (1977) *Science*, **196**, 180–182.
Bruns,R.R., Press,W., Engvall,E., Timpl,R. and Gross,J. (1986) *J. Cell Biol.*, **103**, 393–404.
Carter,W.G. (1982a) *J. Biol. Chem.*, **257**, 13805–13815.
Carter,W.G. (1982b) *J. Biol. Chem.*, **257**, 3249–3257.
Corbi,A.L., Kishimoto,T.K., Miller,L.J. and Springer,T.A. (1988) *J. Biol. Chem.*, **263**, 12403–12411.
Dedhar,S., Ruoslahti,E. and Pierschbacher,M.D. (1987) *J. Cell Biol.*, **104**, 585–593.
Feinberg,A. and Vogelstein,B. (1983) *Anal. Biochem.*, **132**, 6–13.
Fickett,J.W. (1982) *Nucleic Acids Res.*, **10**, 5303–5318.
Furthmayr,H., Wiedemann,H., Timpl,R., Odermatt,E. and Engel,J. (1983) *Biochem. J.*, **211**, 303–311.
Gibson,M.A. and Cleary,E.G. (1985) *J. Biol. Chem.*, **260**, 11149–11159.
Kalafatis,M., Takahashi,Y., Girma,J.-P. and Meyer,D. (1987) *Blood*, **70**, 1577–1583.
Kozak,M. (1987) *Nucleic Acids Res.*, **15**, 8125–8148.
Kyte,J. and Doolittle,R.F. (1982) *J. Mol. Biol.*, **157**, 105–132.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Pareti,F.I., Niiya,K., McPherson,J.M. and Ruggeri,Z.M. (1987) *J. Biol. Chem.*, **262**, 13835–13841.
Pytela,R. (1988) *EMBO J.*, **7**, 1371–1378.
Pytela,R., Pierschbacher,M.D., Ginsberg,M.H., Plow,E.F. and Ruoslahti, E. (1986) *Science*, **231**, 1559–1562.
Roth,G.J., Titani,K., Hoyer,L.W. and Hickey,M.J. (1986) *Biochemistry*, **25**, 8357–8361.
Sakai,L.Y., Keene,D.R. and Engvall,E. (1986) *J. Cell Biol.*, **103**, 2499–2509.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
Schreier,T., Friis,R.R., Winterhalter,K.H. and Trueb,B. (1988) *Biochem. J.*, **253**, 381–386.
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.
Trueb,B., Schaeren-Wiemers,N., Schreier,T. and Winterhalter,K.H. (1989) *J. Biol. Chem.*, **264**, 136–140.
von Heijne,G. (1986) *Nucleic Acids Res.*, **14**, 4683–4690.
Wayner,E.A. and Carter,W.G. (1987) *J. Cell Biol.*, **105**, 1873–1884.

Received on December 13, 1988