FEBS Letters 484 (2000) 164–168 FEBS 24244

EMI, a novel cysteine-rich domain of EMILINs and other extracellular proteins, interacts with the gC1q domains and participates in multimerization

Roberto Doliana[a,1], Simonetta Bot[a,1], Paolo Bonaldo[b], Alfonso Colombatti[a,c,*]

[a] *Divisione di Oncologia Sperimentale 2, CRO-IRCCS, National Cancer Institute, 33081 Aviano, Italy*
[b] *Dipartimento di Istologia e Microbiologia, Università di Padova, 35100 Padova, Italy*
[c] *Dipartimento di Scienze e Tecnologie Biomediche, Università di Udine, 33100 Udine, Italy*

Received 30 September 2000; accepted 12 October 2000

First published online 19 October 2000

Edited by Julio Celis

Abstract The N-terminal cysteine-rich domain (EMI domain) of EMILIN-1 is a new protein domain that is shared with two proteins (multimerin and EMILIN-2) and with four additional database entries. The EMI domains are always located at the N-terminus, have a common gene organization, and belong to proteins that are forming or are compatible with multimer formation. The potential role of the EMI domain in the assembly of EMILIN-1 was investigated by the two-hybrid system. No reporter gene activity was detected when EMI-1 was co-transformed with the C-terminal gC1q-1 domain excluding a head-to-tail multimerization; conversely, a strong interaction was detected when the EMI-1 domain was co-transformed with the gC1q-2 domain of EMILIN-2.

Key words: Extracellular matrix; Elastic fiber; Protein domain; Protein interaction; Two-hybrid system

1. Introduction

EMILIN-1 (formerly gp115 or EMILIN) is an extracellular matrix glycoprotein of the elastic fiber belonging to the C1q/TNF superfamily of proteins [1], and likely involved in elastin deposition [2]. EMILIN-1 forms a fibrillar network in the extracellular matrix of in vitro grown cells and in the matrix of several tissues and it codistributes with elastin in most sites [3–7]. EMILIN-1 is found mainly at the interface between amorphous elastin and microfibrils [2] and its function might be to regulate the formation of the elastic fiber since the addition of EMILIN-1 antibodies to the culture medium of smooth muscle cells greatly perturbs the process of in vitro elastin deposition [2].

The mature EMILIN-1 has a modular organization consisting of an N-terminal cysteine-rich region, an extended region with a high probability of coiled-coil formation, a short collagenic stretch (17 triplets), and a C-terminal globular C1q (gC1q) domain [1]. The gC1q domain is shared with several other ECM constituents including type VIII and type X collagens and, in analogy with several other members of the TNF/C1q superfamily, the native gC1q domain of EMILIN-1 forms trimers [8]. Currently the precise oligomerization state of EMILIN-1 and the exact nature of the forces stabilizing EMILIN-1 assemblies have not been fully identified. Data from gel electrophoresis suggested that S–S bond formation is a contributing factor in EMILIN-1 multimer organization [7,8].

Cysteine-rich domains occur in functionally diverse proteins, including several involved in elastic tissue architecture. Notable is the large number of EGF and calcium binding EGF domains [9,10] that are present as tandem and multiple repeats in fibrillins [11] and in other proteins associated with the elastic fiber [12]. Here, we describe a new superfamily of proteins sharing the novel cysteine-rich domain present in EMILIN-1 and demonstrate its involvement in EMILIN-1 oligomerization.

2. Materials and methods

2.1. Database searches and bioinformatic resources

Sequence searches in the HTGS (high troughput genomic sequences) and EST (expressed sequence tags) GenBank databases were carried out at NCBI web site, using the EMILIN-1 N-terminal region as a query for the BLAST programs with default parameters [13]. The database entries with a significative similarity were used as queries for subsequent rounds of sequence comparison. Gene structure analyses were performed with the GeneScan program [14], located at the Baylor College of Medicine web site.

2.2. RNA extraction and reverse transcriptase/polymerase chain reaction (RT-PCR)

Total cellular RNA was extracted from different cell lines by acid guanidinium thiocyanate–phenol–chloroform [15], and cDNA was synthesized with AMV reverse transcriptase (Promega Corporation, Madison, WI, USA) and random exonucleotides. PCR amplifications were carried out in PCR buffer (10 mM Tris–HCl, pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$), using 2 U of Taq polymerase (Polymed) and 50 pmol each of the following primers: EMI-3 sense 5′-AACTGGTGCCCCTACCCATG-3′, antisense 5′-GGCTTCACCGGGAGCAACTGT-3′; EMI-4 sense 5′-GCCTATGTGGTGCACAGGAATGTG-3′, antisense 5′-GGCTTCACTGGGAAACGCTGC-3′; EMI-5 sense 5′-TCTCTCCACAGGCACTGGTGC-3′, antisense 5′-GGCTTCACCGGGAGCAACTGT-3′; EMI-6 sense 5′-AGGAACTGGTGCTCCTACGTG-3′, antisense 5′-GGACATTCAGGGGTGACCTGT-3′. The amplified products were resolved by agarose gel and directly sequenced to confirm the specificity of the amplified products.

2.3. Construction of plasmids for the two-hybrid system

To study the potential intermolecular interactions of the EMI do-

*Corresponding author. Fax: (39)-434-659 428.
E-mail: acolombatti@ets.it

[1] These two authors contributed equally to the work.

PII: S0014-5793(00)02140-2

main, several constructs encoding LexA fusion proteins were generated by PCR employing 1 ng of EMILIN-pGEM-T as a template [8], 10 pM of each primer (see below), 2.5 U of Taq polymerase (Polymed) and 0.2 mM each of the four dNTPs (Pharmacia Ultrapure, Amersham Pharmacia Biotech). The primers used were the following: sense 5′-CGGAATTCCAGATTGCCCCCCGGCCAGCC-3′ and antisense 5′-CCGCTCGAGTTACGCTGGAGCGGGACTCTCAG-C-3′ for the construct EMI-1 (EMI domain of EMILIN-1, 73 amino acid residues); sense 5′-CGGAATTCCCGGTGTCTTCTCTG-3′ and antisense 5′-CCGCTCGAGTTAGAGGTGGGAAAGGAA-3′ for the construct gC1q-2 (gC1q domain of EMILIN-2, 142 C-terminal amino acid residues). The gC1q-1 (gC1q domain of EMILIN-1, 142 C-terminal amino acid residues) corresponds to the construct yA reported in Mongiat et al. [8]. The underlined bases correspond to appended *Eco*RI and *Xho*I restriction sites. The amplificated fragments were ligated overnight in both pLexA and pB42AD vectors (Clontech Laboratories, Inc.) expressing the binding and the activation domains, respectively, and transformed in *Escherichia coli* competent DH5α strain.

2.4. Yeast transformation and selection of interactors

The *Saccharomyces cerevisiae* strains EGY48 and YM4271 carrying the reporter plasmid p8pop-LacZ were used for the present assays. Growth and manipulation of yeast strains were carried out using the procedures described in the Matchmaker two-hybrid system user manual (Clontech Laboratories, Inc.). Blue colonies usually appeared between 24–112 h. Apart from the standard negative controls, additional controls consisted of EGY48 co-transformed with gC1q-2 and an unrelated clone from a human kidney library.

2.5. Quantitative β-galactosidase assay using ONPG

Overnight cultures in liquid SD selection medium were incubated with SD/Gal/Raf induction medium at 30°C and processed as described (Clontech Laboratories Inc., Yeast Protocol Handbook, Appendix D). To calculate the β-galactosidase units the following formula was applied: $1000\times OD_{420}/t\times V\times OD_{600}$ (t = elapsed time of incubation in min; V = 0.3 ml × concentration factor).

3. Results and discussion

3.1. The EMI domain

An initial NCBI/GenBank database search carried out with the EMILIN-1 N-terminal region as a query for the gapped BLAST-2 program [13] retrieved only one significative entry ($E = 3\times10^{-12}$) corresponding to the N-terminal region of multimerin, a protein involved in complexing the coagulation factor V [16] and also a member of the C1q/TNF superfamily of proteins [16]. Apart from the lack of cysteine number two, the cysteine pattern exhibited identical spacing and several other positions, including aromatic and hydrophobic residues, are conserved between the two proteins, suggesting the existence of a new module that was defined EMI domain. Recently, by means of the two-hybrid system and using as a bait the gC1q-1 domain, a new protein (EMILIN-2) was isolated, whose sequence at the N-terminus revealed a cysteine-rich region possessing the characteristics of the EMI domain. A preliminary description of the general domain organization of EMILIN-2 has been recently published [17], while a full report on this new member will appear elsewhere (Doliana et al., in preparation). Subsequently, database searches using the EMI domain consensus yielded five additional sequences with significative similarity ranging from $E = 10^{-19}$ to 10^{-10} (Fig. 1). Four of these (named provisionally EMI-3, EMI-4, EMI-5 and EMI-6, see the corresponding GenBank accession numbers in Fig. 2), are included in the human and murine databases. The fifth sequence (zf EMI-2) corresponds to a zebra fish EST, and might be the fish hortologue of the EMI domain of EMILIN-2, as it displays a very high similarity to the latter protein ($E = 10^{-17}$). A search against *Drosophila* (http://www.fruitfly.org) and *Caenorhabditis elegans* (http://www.sanger.ac.uk/projects/C_elegans) genomic sequence databases using the EMI domain consensus sequence gave no significant matches, indicating that this domain has been built up during vertebrate evolution as several other domains mainly involved in vascular biology [18].

The EMI domain has seven cysteine residues (C1–C7) located at regular positions. Such cysteine-rich repeat modules, among which the EGF domains are the most studied, have been reported in various other proteins, including several constituents or proteins associated with the elastic fiber [11,12]. An excellent example is fibrillin-1, that presents in its structure at least five different types of cysteine-rich motifs. The EMI domain is rather unique since most of the cysteine-rich domains described to date contain either six or eight cysteine residues. The lengths of the EMI domains range between 71 and 79 residues, and the distances between C1 and C2, C5 and C6, and C6 and C7 are absolutely conserved, whereas minimal gaps are present between C2 and C3, C3 and C4, and C4 and C5 (Fig. 1). In multimerin the EMI domain lacks C2, whereas in EMI-3 C2 is shifted downstream of four residues. Forty-

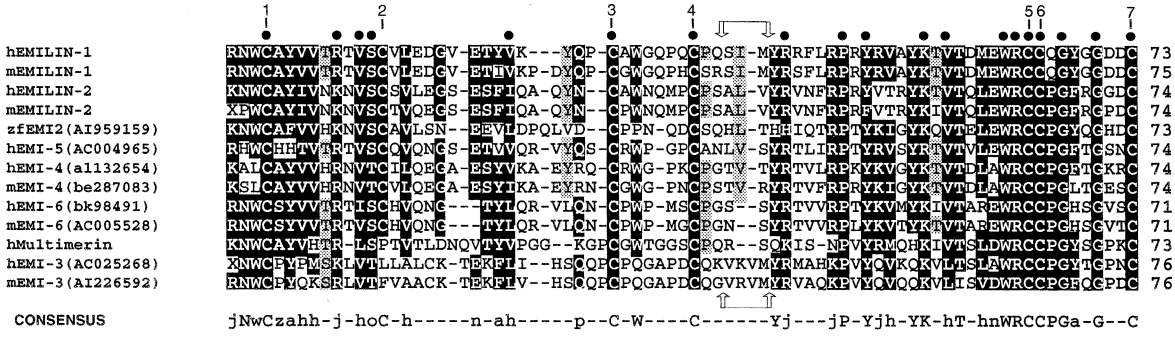

Fig. 1. Multiple sequence alignment of EMI domains performed by CLUSTAL program [22] and refined by eye. Protein names and accession numbers are given on the left. The numbers after the sequences show the length of the domain. Residue positions with conservation of chemical character (see below) in at least 70% of the sequences are highlighted by reverse case, or by gray background (>50%). The positions with 100% conservation are marked also by black dots. The seven characterizing cysteines are numbered and highlighted by arrows. X indicates unidentified residues. The consensus includes residues conserved in the vast majority (>70%) of the aligned sequences: uppercase, identical; a, aromatic (F, Y and W); h, hydrophobic (L, I, V and M); n, acidic (E and D); j, basic (K, R and H); o, polar (S and T); z, tiny (G, A and S); p, amide (N, Q). White arrows delimitate the restricted region where splice junction occurs.

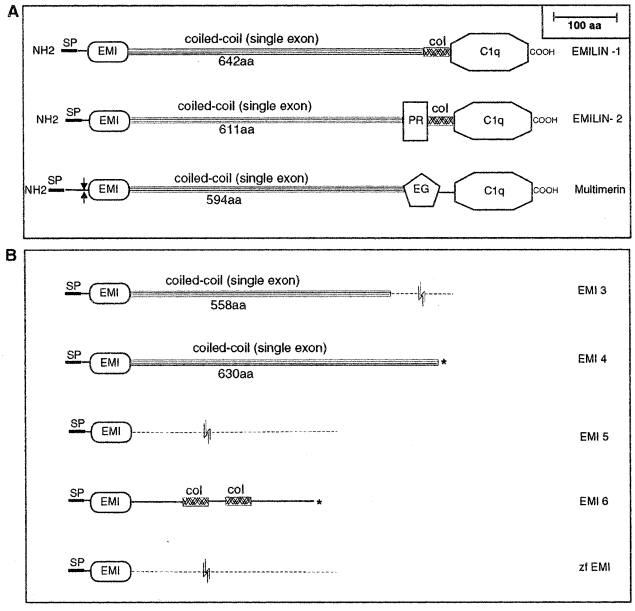

Fig. 2. Modular architecture of the already characterized (A) and candidate (B) members of the EMI domain-containing proteins. Designations are as in Fig. 1. The abbrevations depicted are: aa, amino acids; PR, proline-rich region; col, collagenic stretches; EG, EGF-like domain; C1q, gC1q-like domain; SP, signal peptide; asterisk, stop codon; arrow, pro-peptide cleavage site.

five of the 75 amino acid positions are highly conserved, which allows a strong consensus to be postulated (Fig. 1). In particular, a pattern of WRCCPG(Y/F)xGxxC toward the end of the domain is highly preserved and apparently unique of the EMI domain. Interestingly, in the decapeptide spanning from C5 to C7 the fourth and the seventh residues are always glycines. The consensus sequence of this structure is CxxGxxGxxC, with the distances between residues highly conserved. This sequence arrangement resembles the C-terminal short subdomain of EGF modules that is characterized by a very tight folding consisting of an S-shaped structure with two short stretches of antiparallel β-structures [19]. By analogy with EGF domains, the existence of two subdomains also in the EMI domains could be postulated: a C-terminal subdomain with an S–S bridge between C5 and C7 and a structure resembling an antiparallel β-structure and an N-terminal domain comprised between C1 and C4. In this respect it is relevant to consider that the C-terminal subdomain of EGF is virtually always preceded at position −2 by a cysteine that is bridging with C2; instead, in EMI domains a conserved tryptophan is present in the same position. Thus, apart from the short motif comprised between C5 and C7, there are no other resemblances between EGF and EMI domains: the length of the former type of domain is about 45 residues whereas EMI domains are about 75 residues long; in addition, the N-terminal EMI subdomain has a number of well conserved residues including two tryptophans that are lacking in EGF domains. Nevertheless, the hypothesis that the EMI domains are comprised of two subdomains is corroborated by the gene organization, common to all family members as shown by experimental data for EMI-1 and EMI-2 ([20]; Doliana et al., in preparation), or analysis of clones present in sequence databases for EMI-3, EMI-4, EMI-5, EMI-6, and multimerin: in fact, the domain is always encoded within two exons, the first including C1 to C4, and the second starting right after C4 and including C5 to C7 (Fig. 1).

All the EMI domains are apparently expressed, as indicated by RT-PCR experiments on RNA extracted from the SKUT tumoral cell line (not shown) and by the presence of EST entries for all the family components. EMI-4 is composed of about 700 amino acids (GenBank accession numbers AL132654 and AL1375807) and resembles a truncated form of EMILIN. In addition, for EMI-6 the GeneScan output [13] was verified and corrected using specific primers in RT-PCR; the resulting open reading frame (ORF) encodes for a protein of about 420 amino acids likely belonging to the short chain collagen family as, following the signal peptide and the EMI domain, it has two collagenous sequences (21 and 19 triplets, respectively) and an unrelated C-terminal sequence of about 100 amino acids.

In conclusion, a number of common characteristics for the EMI domain-containing proteins are evident (Fig. 2): (i) the presence of a signal peptide upstream the EMI domains, as assessed by experimental data (for EMILIN-1, EMILIN-2 and multimerin) or by sequence analysis (with the signal IP server at CBS web site), indicates that they are delivered to the extracellular space; (ii) the EMI domains share a common gene organization, being encoded by two exons having similar boundaries (see Fig. 1); this suggests that they might derive from a recent common progenitor; (iii) finally, most of these various components have sequences compatible with the formation of multimers: apart from EMILIN-1 [7,8], EMILIN-2 ([17], Doliana et al., in preparation), and multimerin [16], also EMI-3 and EMI-4 show sequences with a high propensity for coiled-coil formation, as determined by multicoil program analysis [21], and EMI-6 possesses two short collagenic regions (Fig. 2). This kind of organization suggests, even if no structural data are available yet, that three EMI domains might become packed in space.

3.2. *In vivo interactions of the N- and C-terminal (gC1q) domains*

To investigate whether the EMI domain of EMILIN-1 was capable of self-association, the EMI-1 sequence was fused with both the DNA binding domain and with the activation domain of LexA and the interactions were assayed by plating transfectants on selective agar plates. The growth of the yeast transfectants in the incomplete medium and the β-galactosidase activity were then estimated. The EMI-1 domain was shown to interact with itself being able to grow in the incomplete medium and inducing β-galactosidase activity to an extension qualitatively at least comparable to that observed for the gC1q-1/gC1q-1 [8] (Fig. 3A). To check whether EMI-1 interacts with gC1q-1 and might thus contribute to the head-to-tail association between EMILIN-1 homotrimers, the EMI-1 construct was co-transformed with gC1q-1 and no growth was detected even up to 112 h (Fig. 3A). This was also confirmed in liquid cultures using the ONPG as substrate, suggesting that the multimerization of EMILIN-1 [8] did not include a hetero-association between gC1q-1 and EMI-1 domains.

Since the expression pattern of EMILIN-2 was found to be in part overlapping with that of EMILIN-1 (Doliana et al., in preparation), the potential hetero-association between EMI-1

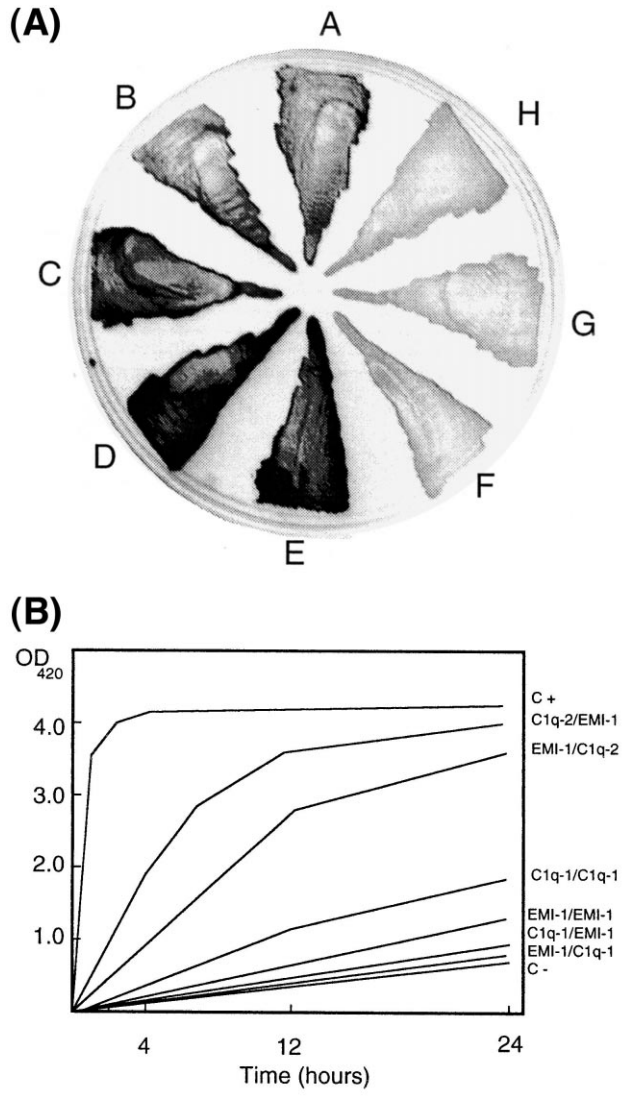

Fig. 3. Interactions of the EMI domain as evaluated by the two-hybrid system. A: Representative colonies of *S. cerevisiae* EGY48 containing the plasmids as indicated have been grown on agar. Yeast colonies containing β-galactosidase activity were identified by the enzymatic conversion of X-gal resulting in blue colored yeast cells. Yeast cells without β-galactosidase activity remained colorless. Each *letter* represents a yeast transformed with different combinations of plasmids with the first partner being cloned in pLexA and the second in pB42AD: A, gC1q-1/gC1q-1; B, EMI-1/EMI-1; C, gC1q-2/EMI-1; D, EMI-1/gC1q-2; E, p53/T antigen; F, gC1q-2/unrelated clone; G, EMI-1/gC1q-1; H, gC1q-1/EMI-1. B: Interactions between EMI and C1q domains were quantitated using a β-galactosidase assay and the ONPG as substrate. C+, p53 fused to the LexA binding domain; C−, an unrelated clone fused to the LexA binding domain.

and gC1q-2 was also investigated: when cells were co-transformed with both EMI-1 and gC1q-2, a strong activation of the reporter gene was detected (Fig. 3A). This interaction was confirmed in the liquid assay: the EMI-1 domain induced a very strong and reciprocal reporter gene activation when combined with gC1q-2 (Fig. 3B). The finding that the data obtained with both reciprocal constructs (EMI-1/gC1q-1 and gC1q-1/EMI-1 as well as EMI-1/gC1q-2 and gC1q-2/EMI-1) were nearly superimposable excludes that differences due to unequal protein expression levels could explain our results.

The initial evidence for EMILIN self-association clearly showed that EMILIN protomers, as several other members of the TNF/C1q superfamily of proteins, associate to form a homotrimer through a relatively stable interaction of their C-terminal gC1q domain [8]. This allows the nucleation of the triple helix and then a further quaternary assembly to higher order polymers via at least intermolecular disulfide bonds [7,8]. This type of supramolecular organization appeared to be unique among the constituents and the associated components of the elastic fiber. With the present data we add another piece of evidence suggesting that EMILIN-1 could interact with EMILIN-2 whereas no binding was demonstrated between the EMI-1 and the gC1q-1 domains. The formation of intermediate hexamers between gC1q-1 domains had also been previously excluded based on the finding that in native gels no bands above the trimer of gC1q-1 were detected [8].

In summary, we identified a new domain for extracellular proteins, characterized by a 7-cysteine characteristic pattern. By both qualitative and quantitative two-hybrid systems we demonstrated an in vivo interaction between EMI-1 and gC1q-2 domain. Recombinant expression of the EMI domains should enable us to perform structural studies that will provide further insights into its spatial organization and functions. Seven out of the eight components of this family have an odd number of cysteine residues and this means that the uncoupled cysteine might be important in disulfide bonding with other proteins and/or for multimerization.

Acknowledgements: This work was supported by grants (to A.C.) from Associazione Italiana per la Ricerca sul Cancro, Telethon (Grant E.704), MURST-Cofin 1998, and Fondo Dipartimentale.

References

[1] Doliana, R., Mongiat, M., Bucciotti, F., Giacomello, E., Deutzmann, R., Volpin, D., Bressan, G.M. and Colombatti, A. (1999) J. Biol. Chem. 274, 16773–16781.
[2] Bressan, G.M., Castellani, I., Colombatti, A., Castellani, I., Marigo, V. and Volpin, D. (1993) J. Cell Biol. 121, 201–212.
[3] Bressan, G.M., Daga-Gordini, D., Colombatti, A. and Volpin, D. (1993) J. Biol. Chem. 258, 13262–13267.
[4] Colombatti, A., Bressan, G.M., Castellani, I. and Volpin, D. (1985) J. Cell Biol. 100, 18–26.
[5] Colombatti, A., Bressan, G.M., Volpin, D. and Castellani, I. (1985) Collagen Rel. Res. 5, 181–191.
[6] Colombatti, A., Poletti, A., Bressan, G.M., Carbone, A. and Volpin, D. (1987) Collagen Rel. Res. 7, 259–275.
[7] Colombatti, A., Bonaldo, P., Volpin, D. and Bressan, G.M. (1988) J. Biol. Chem. 263, 17534–17540.
[8] Mongiat, M., Mungiguerra, G., Bot, S., Mucignat, M.-T., Giacomello, E., Doliana, R. and Colombatti, A. (2000) J. Biol. Chem. 275, 25471–25480.
[9] Cook, R.M., Wilkinson, A.J., Baron, M., Pastore, A., Tappin, M.J., Campbell, I.D., Gregory, H. and Sheard, B. (1987) Nature 327, 339–341.
[10] Engel, J. (1989) FEBS Lett. 251, 1–7.
[11] Pereira, L., D'Alessio, M., Ramirez, F., Lynch, J.R., Sykes, B., Pangilinan, T. and Bonadio, J. (1993) Hum. Mol. Genet. 2, 961–968.
[12] Sinha, S., Nevett, C., Shuttleworth, C.A. and Kielty, C.M. (1998) Matrix Biol. 17, 529–545.
[13] Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Nucleic Acids Res. 25, 3389–3402.
[14] Burge, C. and Karlin, S. (1997) J. Mol. Biol. 268, 78–94.
[15] Chomczynski, P. and Sacchi, N. (1987) Anal. Biochem. 162, 156–159.
[16] Hayward, C.P., Furmaniak-Kazmierczak, E., Cieutat, A.M.,

Moore, J.C., Bainton, D.F., Nesheim, M.E., Kelton, J.G. and Cote, G. (1995) J. Biol. Chem. 270, 19217–19224.
[17] Colombatti, A., Doliana, R., Bot, S., Canton, A., Mongiat, M., Mungiguerra, G., Paron-Cilli, S. and Spessotto, P. (2000) Matrix Biol. 19, 289–301.
[18] Hynes, R.O. and Zhao, Q. (2000) J. Cell Biol. 150, F89–F95.
[19] Appella, E., Weber, I.T. and Blasi, F. (1988) FEBS Lett. 231, 1–4.
[20] Doliana, R., Canton, A., Bucciotti, F., Mongiat, M., Bonaldo, P. and Colombatti, A. (2000) J. Biol. Chem. 275, 785–792.
[21] Wolf, E., Kim, P.S. and Berger, B. (1997) Protein Sci. 6, 1179–1189.
[22] Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G. (1997) Nucleic Acids Res. 25, 4876–4882.