Structural characterization of recombinant domain II of the basement membrane proteoglycan perlecan

Mercedes Costell[a], Takako Sasaki[a], Karlheinz Mann[a], Yoshihiko Yamada[b], Rupert Timpl[a,*]

[a] *Max-Planck-Institut für Biochemie, D-82152 Martinsried, Germany*
[b] *National Institute of Dental Research, NIH, Bethesda, MD 20892, USA*

Received 16 September 1996

Abstract Mouse perlecan domain II (325 residues), consisting of four cysteine-rich LA modules, one IG module and a link region, was obtained in purified form from a stably transfected mammalian cell clone. Rotary shadowing electron microscopy demonstrated a globular domain connected to a short rod-like segment of variable length. This suggested that tandem arrays of LA modules form rod-like elements. Folding into a native structure was indicated by the sharing of immunological epitopes with tissue perlecan, a CD spectrum demonstrating 37% β structure and a limited susceptibility to proteolysis. The domain also showed *N*-glycosylation of a single acceptor site and 7–8 *O*-linked oligosaccharides. The latter were located mainly in the link region within proline-rich sequences.

Key words: Basement membrane; Proteoglycan; Recombinant protein; Rod-like structure

1. Introduction

Perlecan has been identified as the major heparan sulfate proteoglycan of basement membranes and some other extracellular structures and has been shown to bind to integrin cell receptors, laminin, collagen IV, nidogen, fibronectin and several more ligands [1–3]. It has also been implicated in the control of renal filtration through basement membranes and the storage or presentation of cytokines such as basic fibroblast growth factor [4,5]. Mouse perlecan obtained from a tumor consists of an elongated 480 kDa core protein and three glycosaminoglycan chains [6]. Sequence analysis of cDNA clones demonstrated the presence of seven different extracellular protein modules (see [7] for their definition) which are arranged in different clusters referred to as domains I–V [8–10]. The shape and conformation of these domains as well as their specific binding properties are so far largely unknown.

Since perlecan is difficult to purify from tissue extracts and very likely obtained in partially denatured form [6,11], the recombinant production of its individual domains was recently chosen as an alternative approach for structural and functional studies. Such studies were performed for the central domain III of perlecan which consists of alternating arrays of laminin domain IV type (L4) and laminin epidermal growth factor-like (LE) modules [12–14]. Electron microscopical data indicated the presence of globular (L4) domains connected by short rods (LE modules) consistent with X-ray structure data for consecutive laminin LE modules which form rods [15]. The N-terminal perlecan domain I was shown to contain the glycosaminoglycan attachment sites [16]. Domain I was also predicted to contain a novel protein module (SEA) with a mixed α/β conformation [17] which was in fact shown by CD spectroscopy (Costell and Timpl, in preparation).

In the present study we have examined the structure of perlecan domain II which is the connecting segment between domains I and III. Domain II was predicted to contain four LDL receptor class A (LA) modules [7] of about 40 residues including six cysteines, a single immunoglobulin-like (IG) module (100 residues) and a link region of 44 residues [8–10]. The analysis of the recombinant domain II demonstrated that it represents an autonomously folding unit within the perlecan structure and revealed unique spatial features of this protein segment as well as its glycoconjugation.

2. Materials and methods

2.1. Vector construction and recombinant production

Mouse perlecan cDNA clone 54 [8] was used as a template to amplify the sequence encoding domain II (amino acid residues 195–519) by polymerase chain reaction (PCR) with Taq DNA polymerase (Boehringer Mannheim) following the manufacturer's instructions. The primer for the 5′ end was GATCGCTAGCGTTCCCAAGAGTCTGCAC and that for the 3′ end was GATCCTCGAGTCACAGGCAGGAGGCACTGTCTT which introduced, in addition to the coding sequences, a stop codon and single *Nhe*I and *Xho*I restriction sites in order to allow in-frame insertion of the cDNA distal to the BM-40 signal peptide sequence [18]. Several clones, however, showed a C to T mutation at nucleotide position 1976, which would change an Arg to a stop codon, and were repaired by PCR with the 5′ primer CACTGCATCCCCCGAGACTACCTC and 3′ primer GAGGTAGTCTCGGGGGATGCAGTG following an established protocol [19]. The corrected construct was then cloned into the eukaryotic expression vector pRc/CMV (Invitrogen) and its structure verified by DNA sequencing using Sequenase version 2.0 (US Biochemical Co.). Human embryonic kidney 293 cells were cotransfected with the expression vector and plasmid pSV_{2pac}. Stable transfectants were selected with puromycin and serum-free culture medium from regular culture plates was collected as previously described [20].

2.2. Purification of proteins

Recombinant domain II was isolated from 1 liter conditioned serum-free medium on a DEAE cellulose column (2×30 cm) equilibrated in 2 M urea, 0.05 M Tris-HCl pH 8.6 containing protease inhibitors [21]. The protein eluted in a major peak at 0.15–0.2 M NaCl in the linear NaCl gradient (0–0.4 M NaCl, 500 ml). This peak was dialyzed against 0.2 M NH_4HCO_3 and lyophilized to yield almost homogeneous domain II. Perlecan was purified from a tumor basement membrane as described [6] and recombinant perlecan fragments [13,14] from 293 cell culture medium. Proteins were reduced with 0.02 M dithiothreitol in 6 M guanidine and alkylated with iodoacetamide or vinylpyridine following standard protocols.

2.3. Proteolysis and sequence analysis

Digestion of domain II in 0.2 M NH_4HCO_3 with endoproteinases Glu-C (Boehringer Mannheim) and Lys-C (Wako Chemicals, Neuss), trypsin (Worthington) or plasmin (Kabi AB) was carried out at enzyme-substrate ratios of 1:100 at 37°C for different time periods. Major fragments were separated by SDS gel electrophoresis, blotted

*Corresponding author. Fax: (49) (89) 8578 2422.

PII S0014-5793(96)01082-4

onto Immobilon PSQ membranes (Millipore) and used for sequencing. Other aliquots of some digests were reduced and alkylated and separated by reverse-phase chromatography [21] and used to identify glycosylated peptides by hexosamine analysis. Edman degradation of fragments (50–200 pmol) was performed on 473A, 470A and Procise sequencers (Applied Biosystems) following the manufacturer's instructions.

2.4. Analytical and miscellaneous methods

Samples were hydrolyzed with 6 M or 3 M HCl (16 h, 110°C) to determine amino acid and hexosamine compositions, respectively, on a LC 3000 analyzer (Biotronik). SDS gel electrophoresis in 20% polyacrylamide gels followed standard protocols. Circular dichroic spectra were recorded with a Mark IV automatic spectropolarimeter (ISA, Jobin Yvon) at 20°C in quartz cells. Radioimmunoassays [22] and rotary shadowing electron microscopy [23] followed established protocols.

3. Results

The modular structure of perlecan domain II [8] is mainly characterized by four LA modules each with about 40 residues including six invariant cysteines as initially shown for LDL receptor [24]. These modules occur in a tandem array, except modules LA1 and LA2 which are separated by a 44 residue link region, followed by a single immunoglobulin-like module IG1 and an extra disulfide-bonded loop (E) at the C-terminus (Fig. 1a). An expression vector for domain II of mouse perlecan (position 195–519) with a foreign signal peptide was prepared and several stably transfected kidney cell clones were obtained which produced and secreted recombinant fragment II. Purification to >95% homogeneity could be achieved by a single chromatographic step on DEAE cellulose with high yields (15–20 mg fragment II per liter culture medium). Edman degradation of fragment II demonstrated the expected N-terminal sequence APLAFPRV with APLA being derived from the signal peptide region [8,18]. SDS gel electrophoresis showed a single band of about 45 kDa for non-reduced fragment II (Fig. 2), in agreement with a molecular mass of 36 kDa calculated from the sequence assuming some additional glycosylation. Glycosylation of the recombinant protein was confirmed by hexosamine analysis showing 7.3±1.2 residues of glucosamine and 7.6±1.0 residues of galactosamine. This indicates full occupation of the single *N*-glycosylation site in the LA3 module (Fig. 1a) and about 7 *O*-glycosylation sites in fragment II.

Rotary shadowing electron microscopy of recombinant fragment II demonstrated a rather uniform set of particles with an average length of 18±4 nm (Fig. 3). Typical for this structure was a small globular domain connected to a thin stalk of 10–15 nm length. A distinct folding of the fragment was also shown by its CD spectrum which had a distinct minimum at 198 nm (Fig. 4). This indicated a 37% content of β structure with only low or insignificant amounts of α helix [25]. These observations also suggested an extensive and probably correct disulfide bonding among the 28 cysteines present in fragment II. Disulfide bonding was indicated by a distinct decrease in electrophoretic mobility observed after reduction of fragment II (Fig. 2). Further evidence was obtained by radioimmuno-inhibition assay used for epitope analysis of a polyclonal rabbit antiserum against fragment II (Fig. 5). This demonstrated equally high inhibitory capacities for non-reduced fragment II and tissue-derived perlecan but a more than 10000-fold loss of reduced and alkylated fragment II. Furthermore, other recombinant perlecan fragments such as III-3 [13] showed no inhibition demonstrating the presence of unique conformational antibody binding epitopes on fragment II.

Proteolysis and sequence analysis were performed to examine disulfide bonding and the presence of flexible hinge regions and oligosaccharide substitutions. Digestions with trypsin, endoproteinases Lys-C or Glu-C produced similar but not identical cleavage patterns of fragment II including components of about 40, 22 and 4/5 kDa when analyzed by electrophoresis under non-reducing conditions (Fig. 2). No reduction in mobility was, however, observed after a 24 h treatment with plasmin. Yet, all four digests showed a more distinct increase in mobility, including that with plasmin (components of 20–43 kDa), of individual components when examined under reducing conditions (data not shown), indicating that cleavage occurred in intermodular segments as well as within disulfide-bonded loops of individual modules. This could be confirmed by sequence analysis of the major fragments already released without reduction (Table 1, Fig. 1b). The data for the 40 kDa band demonstrated a release after cleavage within the

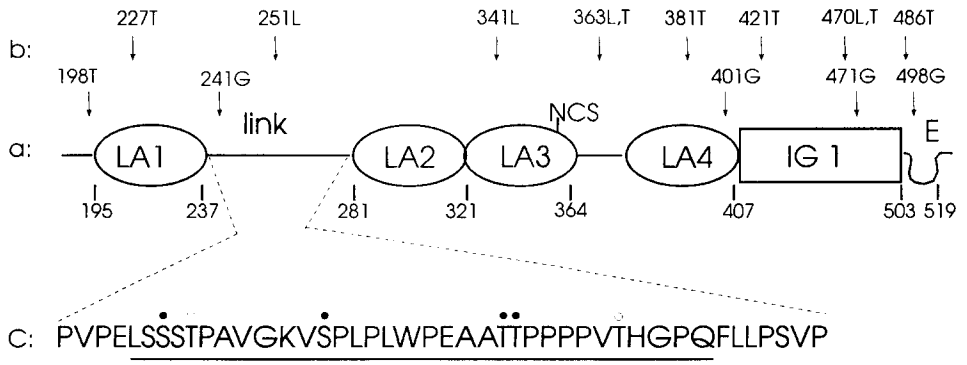

Fig. 1. Module arrangement of perlecan domain II and identification of major protease sensitive bonds and oligosaccharide attachment sites. a: Modules include four LDL receptor class A repeats (LA), one immunoglobulin superfamily repeat (IG), an extra disulfide-bonded loop (E) and a few link regions (straight lines). Numbers underneath indicate their approximate borders by sequence positions [8]. NCS denotes the single *N*-glycosylation site. b: Approximate positions of major cleavage sites by trypsin (T), endoproteinase Glu-C (G) and endoproteinase Lys-C (L) identified by the position numer of the newly generated N-terminus (see Table 1). c: Predicted cDNA sequence of the major link region [8] with the portion confirmed by Edman degradation being underlined. A gap in the sequence or a strongly reduced yield of the respective PTH-amino acid was interpreted to indicate full (●) or partial (○) *O*-glycosylation, respectively.

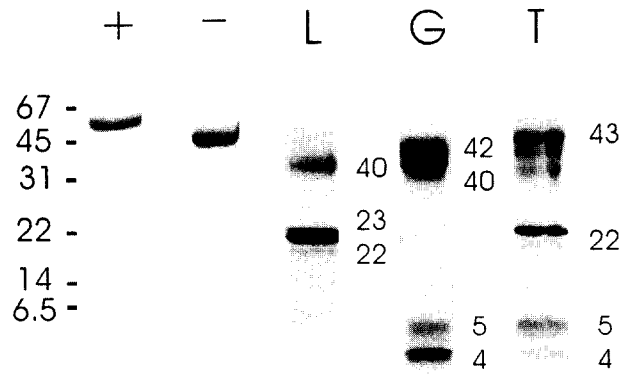

Fig. 2. Electrophoretic analysis of recombinant perlecan domain II and of several proteolytic digests. The left two lanes show intact fragment II after (+) and prior (−) to reduction. Numbers at the left margin denote the position of calibrating proteins in kDa. Proteolytic digests of fragment II were obtained with endoproteinases Lys-C (L) and Glu-C (G) each for 4 h and with trypsin (T) for 1 h and separated under non-reducing conditions. Major fragments subjected to sequencing are identified by their molecular masses (in kDa).

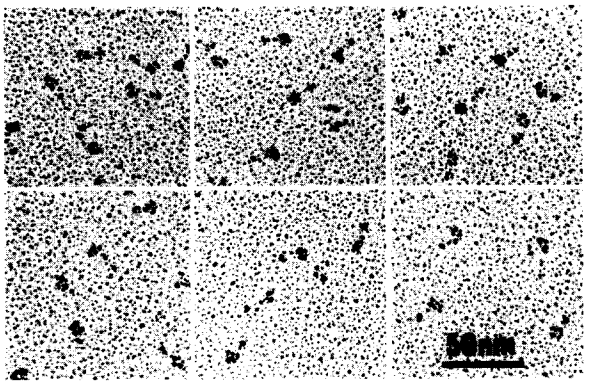

Fig. 3. Electron micrograph of rotary shadowed fragment II. Note the frequent appearance of a globular structure connected to a short stalk.

link between LA1 and LA2 modules (G40,42, Table 1) or demonstrated the N-terminus (L40, T43, Table 1) indicating here shortening at the C-terminus. All fragments contained in addition internal loop cleavage sites within the LA1, LA3, LA4 and/or the IG1 module. The 22 kDa components were apparently released by a cleavage between modules LA3 and LA4 with additional cleavages within the LA4 or IG1 module. The 4/5 kDa double bands corresponded to either the LA1 or the extra loop (E) modules or portions of the IG1 module.

Initial sequence analysis of fragments starting in the predicted large link region between the LA1 and LA2 modules (L26,32; G40,42; see Table 1) showed in addition the interesting feature of a gap in the sequence at positions that should be occupied by Ser or Thr [8] indicating the possible modification by *O*-glycosylation. Reduced and alkylated fragments were therefore further purified by reversed phase chromatography and revealed only two peaks which contained more than 80% of the total galactosamine recovered (data not shown). These fragments allowed a substantial sequence analysis of the link region demonstrating full substitution for Ser^{243}, Ser^{252}, Thr^{262} and Thr^{263} and partial substitution ($>50\%$) for Thr^{245} and Thr^{269} (Fig. 1c) estimated from the yields of derivatized amino acids.

4. Discussion

The data reported in this study demonstrate by recombinant production that the predicted domain II of the basement membrane proteoglycan perlecan represents an autonomous folding unit within its large core protein. This autonomy may even exist at the level of individual modules which, however, remains to be shown. Folding of recombinant fragment II into a native structure was indicated by electron microscopy, by the presence of a distinct secondary structure as

Table 1
Amino-terminal sequences of major fragments produced from perlecan domain II by cleavage with endoproteinases Glu-C (G) and Lys-C (L) and by trypsin (T)

	Protease, fragment size (kDa)	Sequence	Cleaved peptide bond and position	Localization in modules
L	40	APLAFPRV	–	N-terminus
		LWRXD	KL,340/341	LA3
	22,23	QPGEVX	KL,362/363	border LA3/4
		EADQG	KE,469/470	IG1
	26,32[a]	APLAFPRV	–	N-terminus
		VXPLPLW	KV,250/251	link
G	40,42	LSXSTPAV	EL,240/241	link
		FGXMPP	EF,400/401	LA4
		ADQGA	EA,471/472	IG1
	4–5	APLAFRV	–	N-terminus
		LVPQR	EL,497/498	border IG1/E
T	43	VVXTET	RV,197/198	front LA1
		DMSDELN	RD,226/227	LA1
		XIPASFH	RC,380/381	LA4
	22	QPGEV	KL,362/363	border LA3/4
		XIPAS	RC,380/381	LA4
	4–5	GQTVT	RG,420/421	IG1
		EADQG	KE,470/471	IG1
		GMVFG	RG,485/486	border IG1/E

[a]Minor component by protein staining.
Components were separated by electrophoresis under non-reducing conditions prior to sequencing. Compare Fig. 2 for the identification of fragments and Fig. 1 for modules.

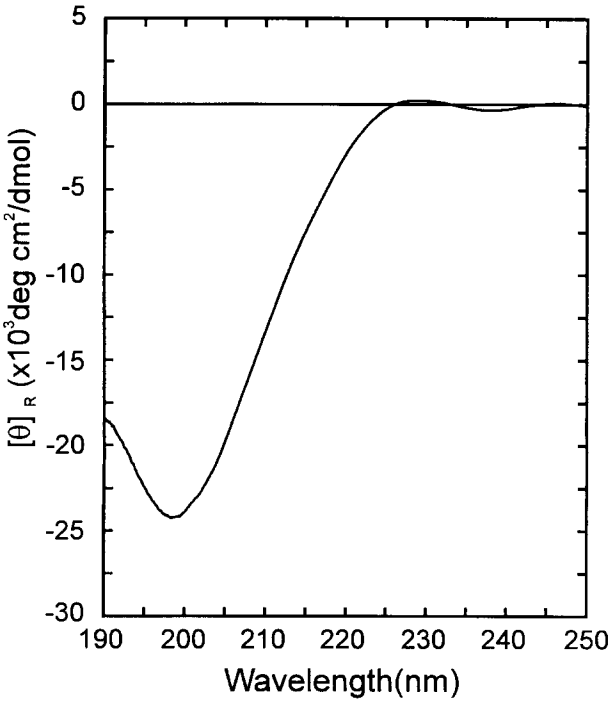

Fig. 4. Circular dichroism spectrum of recombinant perlecan fragment II.

shown by CD spectroscopy and by the complete cross-reaction with tissue-derived perlecan in quantitative immunochemical assays. Furthermore an extensive disulfide bonding within fragment II was demonstrated by proteolysis and the complete loss of immunological epitopes upon reduction and alkylation.

A globular domain and a thin rod represent the topological features of fragment II. These two structural elements are also typical for the electron microscopical appearance of the core protein of perlecan [6] indicating that domain II could contribute about 20% of the total length of the core. It is very likely that the globular domain corresponds to the IG1 module and accounts for most of the β structure, since these are typical features of similar modules found in immunoglobulins [26]. This suggests that the tandem array of four LA modules contributes the rod-like element of fragment II which was so far not known before. Such modules obviously must have a compact structure due to the three overlapping disulfide bridges as shown for a single recombinant LA module of LDL receptor [27]. Similar compact structures are known to exist in the laminin-type LE modules and EGF-like modules which are of comparable size but have a different disulfide bond connectivity [15,28]. Also these latter modules have been shown to form rod-like structures if they occur in tandem arrays. Whether tandem arrays of LA modules form rod-like structures will require further recombinant analysis of perlecan domain II as well as of other proteins showing similar arrangements. The only other protein known so far is the LDL receptor which contains a consecutive array of seven LA module at its N-terminal region [24].

Proteolysis was used as a probe to identify connecting intermodular segments within perlecan domain II. This demonstrated cleavage of a 44 residue link region which separates the LA1 and LA2 modules and is particularly rich in Pro, Thr and Ser (Fig. 1c). Some shorter protease-sensitive connecting segments apparently exists between modules LA3 and LA4 and possibly between LA4 and IG1. Yet significant cleavage was also observed within disulfide-bonded loops of IG1 and three LA modules (Fig. 1b) which indicates the existence of exposed loop structures. These observations are supported by a similar trypsin sensitivity of the LA1 module of the LDL receptor [27] which shows that this property is a general feature of LA modules. The link region between the LA1 and LA2 modules of perlecan domain II was also shown to harbor six *O*-glycosylation sites consistent with the presence of 7–8 galactosamines and particular features of its sequence. This indicates a short mucin-like segment in domain II which could be responsible for the size variation observed for the rod-like segment. It was also of interest that all except one of the substituted Ser and Thr had a Pro in the +3 position which fulfils one of the requirements postulated for *O*-glycosylation [29].

The LA module region of LDL receptor was previously shown to be important for the binding of extracellular ligands. A mutational analysis demonstrated that LDL binding through ApoB-100 required the consecutive array of five LA modules while only the LA5 module was necessary for promoting VLDL binding through ApoB [30]. Site-directed mutagenesis further identified a conserved Ile and Asp as crucial residues within the binding epitopes. Both residues are invariant, except for a few conservative substitutions, in the LA modules of perlecan [8–10]. Nidogen [1] and fibulin-2 [31] are the only extracellular binding proteins of the perlecan core protein that have been identified so far. Preliminary evidence was obtained that nidogen binds to recombinant fragment II although less strongly than to perlecan (M. Costell and R. Timpl, unpublished). Further studies with the recombinant fragment will therefore be necessary to establish whether do-

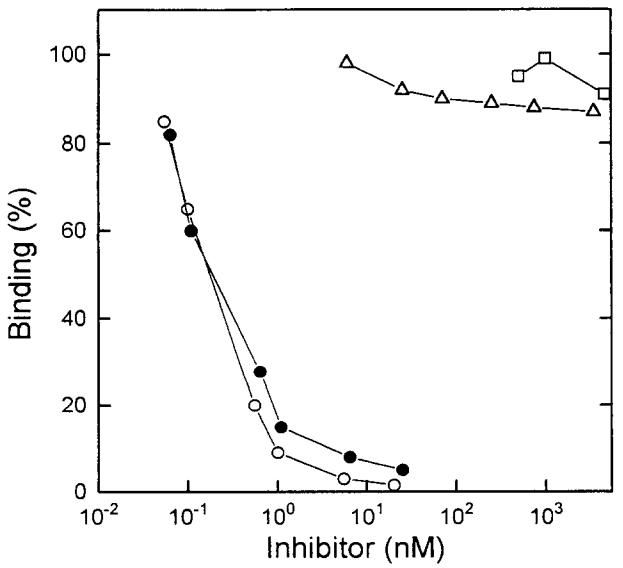

Fig. 5. Immunochemical analysis of fragment II by radio-immunoinhibition assay. The assay system consisted of ^{125}I-labelled fragment II and a rabbit antiserum against fragment II. Inhibitors used were perlecan (●), perlecan fragment II (○), reduced and alkylated fragment II (△) and perlecan fragment III-3 (□).

main II has distinct binding properties or just serves as a spacing element within the elongated core protein of perlecan.

Acknowledgements: The study was supported by a research training fellowship from the European Community to M.C. (ERBCHBIC-T941267). We are grateful for the skilled technical assistance of Mrs. Hanna Wiedemann and Mischa Reiter.

References

[1] Battaglia, C., Mayer, U., Aumailley, M. and Timpl, R. (1992) Eur. J. Biochem. 208, 359–366.
[2] Timpl, R. (1993) Experientia 49, 417–428.
[3] Iozzo, R., Cohen, I.R., Grässel, S. and Murdoch, A.D. (1994) Biochem. J. 302, 625–639.
[4] Aviezer, D., Hecht, D., Safran, M., Eisinger, M., David, G. and Yayon, A. (1994) Cell 79, 1005–1013.
[5] Whitelock, J.M., Murdoch, A.D., Iozzo, R.V. and Underwood, P.A. (1996) J. Biol. Chem. 271, 10079–10086.
[6] Paulsson, M., Yurchenco, P.D., Ruben, G.C., Engel, J. and Timpl, R. (1987) J. Mol. Biol. 197, 297–313.
[7] Bairoch, A. (1996) SWISS-PROT Protein Sequence Data Bank, Release 33.0.
[8] Noonan, D.M., Fulle, A., Valente, P., Cai, S., Horigan, E., Sasaki, M., Yamada, Y. and Hassell, J.R. (1991) J. Biol. Chem. 266, 22939–22947.
[9] Kallunki, P. and Tryggvason, K. (1992) J. Cell Biol. 116, 559–571.
[10] Murdoch, A.D., Dodge, G.R., Cohen, I., Tuan, R.S. and Iozzo, R.V. (1992) J. Biol. Chem. 267, 8544–8557.
[11] Hassell, J.R., Leyshon, W.C., Ledbetter, S.R., Tyree, B., Suzuki, S., Kato, M., Kimata, K. and Kleinman, H.K. (1985) J. Biol. Chem. 260, 8098–8105.
[12] Chakravarti, S., Horchar, T., Jefferson, B., Laurie, G.W. and Hassell, J.R. (1995) J. Biol. Chem. 270, 404–409.
[13] Schulze, B., Mann, K., Battistutta, R., Wiedemann, H. and Timpl, R. (1995) Eur. J. Biochem. 231, 551–556.
[14] Schulze, B., Sasaki, T., Costell, M., Mann, K. and Timpl, R. (1996) Matrix Biol. (in press).
[15] Stetefeld, J., Mayer, U., Timpl, R. and Huber, R. (1996) J. Mol. Biol. 257, 644–657.
[16] Kokenyesi, R. and Silbert, J.E. (1995) Biochem. Biophys. Res. Commun. 211, 262–267.
[17] Bork, P. and Patthy, L. (1995) Protein Sci. 4, 1421–1425.
[18] Mayer, U., Nischt, R., Pöschl, E., Mann, K., Fukuda, K., Gerl, M., Yamada, Y. and Timpl, R. (1993) EMBO J. 12, 1879–1885.
[19] Kohfeld, E., Göhring, W., Mayer, U., Zweckstetter, M., Holak, T.A., Chu, M.-L. and Timpl, R. (1996) Eur. J. Biochem. 238, 333–340.
[20] Nischt, R., Pottgiesser, J., Krieg, T., Mayer, U., Aumailley, M. and Timpl, R. (1991) Eur. J. Biochem. 200, 529–536.
[21] Mayer, U., Aumailley, M., Mann, K., Timpl, R. and Engel, J. (1991) Eur. J. Biochem. 198, 141–150.
[22] Timpl, R. (1982) Methods Enzymol. 82, 472–498.
[23] Engel, J. and Furthmayr, H. (1987) Methods Enzymol. 145, 3–78.
[24] Südhof, T.C., Goldstein, J.L., Brown, M.S. and Russell, D.W. (1985) Science 228, 815–822.
[25] Provencher, S.W. and Glöckner, J. (1981) Biochemistry 20, 33–37.
[26] Branden, C. and Tooze, J. (1991) Introduction to Protein Structure. Garland, New York, pp. 179–199.
[27] Bieri, S., Djordjevic, J.T., Daly, N.L., Smith, R. and Kroon, P.A. (1995) Biochemistry 34, 13059–13065.
[28] Campbell, I.D. and Bork, P. (1993) Curr. Opin. Struct. Biol. 3, 385–392.
[29] Wilson, I.B.H., Gavel, Y. and von Heijne, G. (1991) Biochem. J. 275, 529–534.
[30] Russell, D.W., Brown, M.S. and Goldstein, J.L. (1989) J. Biol. Chem. 264, 21682–21688.
[31] Sasaki, T., Göhring, W., Pan, T.-C., Chu, M.-L. and Timpl, R. (1995) J. Mol. Biol. 254, 892–899.