Crystal structure of the angiogenesis inhibitor endostatin at 1.5 Å resolution

Erhard Hohenester[1], Takako Sasaki[2], Björn R.Olsen[3] and Rupert Timpl[2]

Department of Crystallography, Birkbeck College, London WC1E 7HX, UK, [2]Max-Planck-Institut für Biochemie, D-82152 Martinsried, Germany and [3]Department of Cell Biology, Harvard Medical School, Boston, MA 02115, USA

[1]Corresponding author
e-mail: e.hohenester@cryst.bbk.ac.uk

A number of extracellular proteins contain cryptic inhibitors of angiogenesis. Endostatin is a 20 kDa C-terminal proteolytic fragment of collagen XVIII that potently inhibits endothelial cell proliferation and angiogenesis. Therapy of experimental cancer with endostatin leads to tumour dormancy and does not induce resistance. We have expressed recombinant mouse endostatin and determined its crystal structure at 1.5 Å resolution. The structure reveals a compact fold distantly related to the C-type lectin carbohydrate recognition domain and the hyaluronan-binding Link module. The high affinity of endostatin for heparin is explained by the presence of an extensive basic patch formed by 11 arginine residues. Endostatin may inhibit angiogenesis by binding to the heparan sulphate proteoglycans involved in growth factor signalling.
Keywords: angiogenesis inhibitor/collagen/crystal structure/heparin binding

Introduction

In the adult mammal, new blood vessels are formed by angiogenesis, the sprouting of new capillaries from existing vasculature (Bussolino *et al.*, 1997; Risau, 1997). The endothelium lining a mature vessel is normally quiescent and angiogenesis in the adult generally is associated with pathological situations, such as wound healing and tumour growth. Angiogenesis is a complex multi-stage process that involves, in approximate temporal order, the proteolytic degradation of the basement membrane, loss of endothelial cell adhesion, proliferation and migration of endothelial cells into the surrounding stroma, and finally re-adhesion of endothelial cells to form the lumen of the new capillary tube. Ongoing angiogenesis is essential for the rapid growth of solid tumours and it appears that successful tumours actively influence the 'angiogenic switch' to sustain continuous cell proliferation. The strict dependence on recruiting new blood vessels from the host makes tumours vulnerable to anti-angiogenic therapy (Hanahan and Folkman, 1996).

A number of inducers of angiogenesis have been identified (Hanahan and Folkman, 1996; Beck and D'Amore, 1997; Bussolino *et al.*, 1997). Of these, acidic and basic fibroblast growth factor (aFGF and bFGF, respectively) and vascular endothelial growth factors (VEGF) are the most widely expressed in normal adult organs. These growth factors are recognized on endothelial cells by transmembrane tyrosine kinase receptors coupled with the intracellular signal transduction pathways. Cell surface heparan sulphate proteoglycans play a key role in bFGF signalling, either by assisting the association of bFGF with its receptor or by promoting bFGF oligomerization (Yayon *et al.*, 1991; Friesel and Maciaq, 1995; Herr *et al.*, 1997; Moy *et al.*, 1997).

Given the requirement for tight control of angiogenesis, it is not surprising that there exist a number of endogenous angiogenesis inhibitors, and that imbalance in their levels is associated with tumorigenesis (Hanahan and Folkman, 1996). One well-established example is the secreted glycoprotein thrombospondin-1 (TSP-1), which is regulated by the p53 tumour suppressor protein (Good *et al.*, 1990; Dameron *et al.*, 1994). Both full-length TSP-1 and specific fragments of it inhibit angiogenesis (Good *et al.*, 1990; Tolsma *et al.*, 1993). In contrast, some of the most potent angiogenic inhibitors are fragments derived from abundant extracellular proteins that themselves do not regulate angiogenesis. A 29 kDa fragment of fibronectin (Homandberg *et al.*, 1985), a 16 kDa fragment of prolactin (Ferrara *et al.*, 1991; Clapp *et al.*, 1993) and a 38 kDa fragment of plasminogen termed angiostatin (O'Reilly *et al.*, 1994) are all inhibitors of endothelial cell proliferation, whereas their parent proteins are not. The storage of a class of angiogenesis inhibitors as cryptic fragments of abundant proteins associated with the vascular system appears to be a recurring theme in the regulation of angiogenesis (Hanahan and Folkman, 1996; Sage, 1997). Many angiogenesis inhibitors bind with high affinity to heparin, suggesting antagonization of bFGF signalling as a possible mechanism of action (Brown and Parish, 1994).

Recently, a new angiogenesis inhibitor was identified that turned out to be a 20 kDa C-terminal fragment of collagen XVIII. This heparin-binding fragment, termed endostatin, specifically inhibits endothelial cell proliferation and potently inhibits angiogenesis and tumour growth (O'Reilly *et al.*, 1997). Cycled therapy with recombinant endostatin reduced several experimental tumours, including Lewis lung carcinoma, to a dormant state and did not induce resistance. Even more remarkably, after a few cycles of treatment dormancy persisted even when therapy was discontinued (Boehm *et al.*, 1997).

The $\alpha1$(XVIII) collagen is an unusual collagen characterized by 10 domains of triple-helical collagenous repeats separated by non-triple-helical repeats (Oh *et al.*, 1994a; Rehn and Pihlajaniemi, 1994). It is expressed in a tissue-specific manner as three alternative splice variants and is localized mainly in perivascular basement membrane zones (Muragaki *et al.*, 1995; Rehn and Pihlajaniemi, 1995). The last six of the triple-helical repeats of $\alpha1$(XVIII)
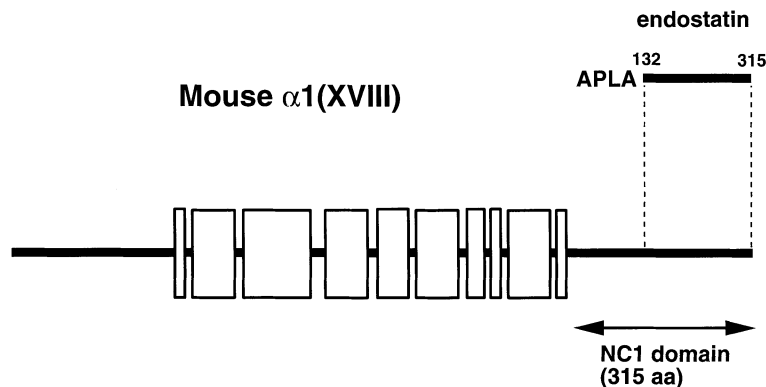
endostatin

Mouse α1(XVIII)

APLA

132   315

NC1 domain
(315 aa)

Fig. 1. Domain structure of mouse α1(XVIII) collagen. Triple-helical domains are indicated by rectangular boxes; non-triple-helical regions are indicated by heavy lines. The region in the C-terminal NC1 domain that corresponds to the proteolytic fragment endostatin is indicated. The numbering scheme is based on the position in the NC1 domain (see text). The recombinant endostatin used in this work contains the APLA sequence at the N-terminus; this sequence is a result of the cloning process and is fortuitously identical to the α1(XVIII) collagen NC1 sequence at Leu130 and Ala131.

collagen are almost identical in size to those of α1(XV) collagen (Myers *et al.*, 1992), and the name multiplexins (for <u>multiple</u> triple heli<u>x</u> domains and <u>in</u>terruptions) has been given to this new collagen family (Oh *et al.*, 1994a). The α1(XVIII) collagen contains a non-collagenous C-terminal domain (NC1) of ~300 residues; the angiogenic inhibitor endostatin corresponds to the last 184 amino acid residues of the NC1 domain (Figure 1). Interestingly, this exactly matches the conserved region between the α1(XVIII) and α1(XV) C-terminal domains (Oh *et al.*, 1994a).

Given the considerable potential of endostatin as an anti-cancer therapeutic, we would like to understand the molecular mechanisms by which this protein inhibits endothelial cell proliferation and angiogenesis. To help design experiments addressing these questions, we have solved the crystal structure of endostatin at 1.5 Å resolution. A large basic surface area is proposed to be the heparin-binding site of endostatin. Thus, endostatin may exert its antiproliferative effect by competing with bFGF for binding to cell surface heparan sulphate proteoglycans, which could disrupt the mitogenic growth factor signal.

Results and discussion

Protein expression and structure determination
Mouse endostatin was expressed at high levels as soluble protein in human embryonic kidney cells and was found to potently inhibit the bFGF-induced proliferation of human umbilical vein endothelial cells, with an IC_{50} of ~100 ng/ml (unpublished data). The secreted protein spans the 184 C-terminal amino acid residues of mouse α1(XVIII) collagen and, additionally, contains the N-terminal sequence APLA (Figure 1). To avoid ambiguities caused by the N-terminal splicing of α1(XVIII) collagen (Muragaki *et al.*, 1995; Rehn and Pihlajaniemi, 1995), we decided to number the endostatin sequence relative to its position in the NC1 domain, starting at His132.

The structure of endostatin was solved using crystals grown from ammonium phosphate at pH 5. Phases to 2.2 Å resolution were obtained by the multiple isomorphous replacement method and anomalous scattering (MIRAS)

from three heavy-atom derivatives (Table I). The resulting electron density map was of high quality, allowing the polypeptide chain to be traced without difficulty. The final structure, refined at 1.5 Å resolution (Figure 2; Table II), consists of residues Gln138 to Phe309 and a total of 83 water molecules. The first ten and last six residues are not visible and are presumed to be disordered.

Description of the overall structure
Endostatin folds into a single globular domain of approximate dimensions 35×30×30 Å (Figure 3). The structure is composed predominantly of a β-sheet and loops, but also contains two α-helices, one of them short. A total of 40% of the amino acid residues adopt an extended main chain conformation, but due to many irregularities such as kinks and bulges, only 25% are actually contained in uninterrupted β-strands spanning more than two residues. The fold of endostatin is intricate and is best described with reference to a schematic representation (Figure 3C). The most prominent feature of the structure is a highly twisted mixed β-sheet composed of seven strands (E, F, A, P, J, M and O). An α-helix of 14 residues, α1, packs against one face of this sheet, whereas the other face is covered by elaborate loop structures involving a short stretch of antiparallel β-sheet (strands G and N) and a second shorter α-helix, α2. At the C-terminus of strand A, the polypeptide chain bulges around a water molecule before re-establishing hydrogen bonding with the N-terminus of strand P for another two residues. The segment following α1 forms the short strand B, then kinks at Phe180 and leads into the β-hairpin C–D. Strands B, J and P form a triangular structure, at the centre of which a water molecule hydrogen bonds with the peptide carbonyls of Phe180 and Trp251 and the amide nitrogen of Leu303. This water molecule is deeply buried in the hydrophobic core and represents an integral part of the endostatin structure. Apart from the C–D β-hairpin, there are two additional classical β-hairpins, one extending the central β-sheet at strand A (strands H and I), the other following strand J after a kink at Gly253 (strands K and L). The overall arrangement of β-strands in the endostatin structure can be described as a rather irregular β-barrel propped open on one side by α2.

Table I. Data collection and phasing statistics

	Native I	Native II	UO$_2$SO$_4$	K$_2$Pt(CN)$_4$	NaAu(CN)$_2$
Data collection					
Resolution range (Å)	15–2.0	25–1.5	15–2.2	14–2.2	15–2.2
Observations	57717	159011	40029	33343	40158
Unique reflections	10927	26471	8240	8259	8451
Completeness (%)	96.3	99.3	95.8	96.5	98.4
R_{merge}[a]	0.055	0.075	0.074	0.042	0.057
Phasing statistics					
Resolution range (Å)			15–2.2	15–2.2	15–3.5
R_{deriv}[b]			0.187	0.133	0.207
Heavy-atom sites			3	4	2
R_{Cullis}[c]			0.78/0.71	0.77/0.74	0.87/0.85
Phasing power[d]			1.4/1.0	1.3/1.0	0.8/0.7

[a]$R_{merge} = \Sigma_h\Sigma_i|I_i(h) - <I(h)>|/\Sigma_h\Sigma_iI_i(h)$, where $I_i(h)$ is the i-th measurement of reflection h and $<I(h)>$ is the weighted mean of all measurements of h.
[b]$R_{deriv} = \Sigma_h||F_{deriv}(h)| - |F_{nat}(h)||/\Sigma_hF_{nat}(h)$, where F_{nat} and F_{deriv} are the native and derivative structure factor amplitudes, respectively.
[c]$R_{Cullis} = \Sigma_h|F_{deriv}(h) - F_{nat}(h)| - |F_H|/\Sigma_h|F_{deriv}(h) - F_{nat}(h)|$, where F_H is the calculated heavy-atom structure factor. Pairs of values are for acentric and centric reflections, respectively.
[d]The phasing power is defined as (r.m.s. F_H/r.m.s. lack-of-closure) summed over the reflections used in the heavy-atom refinement. Pairs of values are for acentric and centric reflections, respectively.
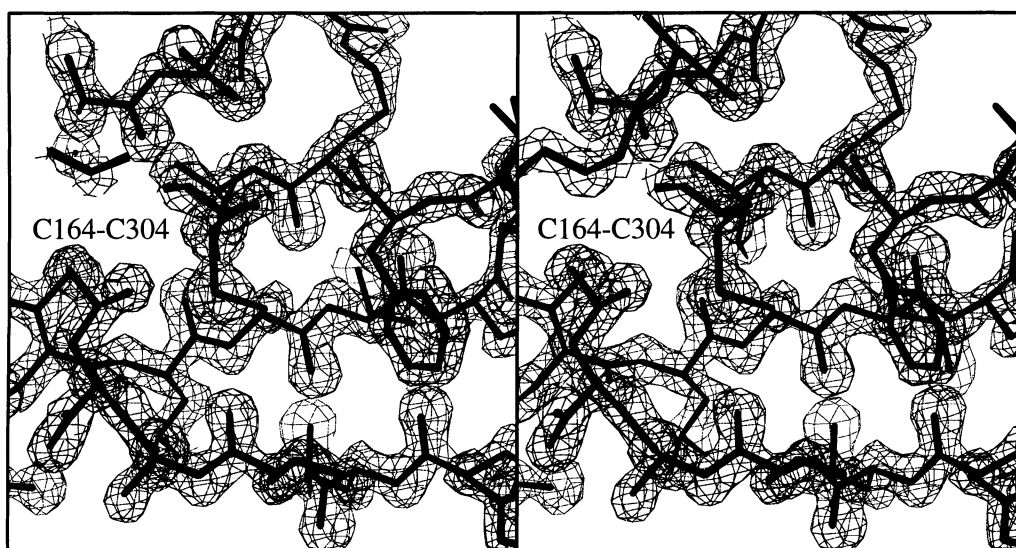

Fig. 2. Stereo view of the final $2F_{obs}-F_{calc}$ map at 1.5 Å resolution. The region around β-strand P is shown, with the disulphide bond between Cys164 and Cys304 in the centre. The map is contoured at the 1.5 σ level and is shown with the refined model superimposed. The figure was made with BOBSCRIPT (R.Esnouf, personal communication), a modified version of MOLSCRIPT (Kraulis, 1991).

Endostatin contains two disulphide bridges in a nested pattern, linking Cys164 with Cys304 and Cys266 with Cys296. The former disulphide bridge connects α1 to the central β-sheet, and the latter circularizes a twisted loop containing strands M, N and O.

Despite the disjointed fold and the large fraction of irregular loop structures, endostatin is a compact molecule. All surface loops pack tightly against the body of the structure and many of them contribute to a large hydrophobic core. This core is divided into two regions of different size by strand J. The smaller core is centred around Trp251 and is built up mainly from amino acid side chains contributed by strands J, M, O and P. A much more extensive hydrophobic core fills the large concave face of the central β-sheet and this is also the region where most of the longer surface loops are found.

Strands A and P of endostatin are situated next to each other, engaged in antiparallel hydrogen bonding. As a result, the first and last residue defined by the electron density (Gln138 and Phe309, respectively) are close in space. In the intact NC1 domain of collagen XVIII, endostatin is preceded by 131 residues. This N-terminal portion of unknown structure is likely to interact with endostatin near Phe309 and this interaction may well lead to an ordering of the C-terminus of endostatin, which significantly contains two large hydrophobic residues, Met310 and Phe313. We note that the side chains of Phe162 and Phe165, located next to each other on α1, are fully exposed to solvent (Figure 5A). In the crystal lattice they are covered by a hydrophobic packing contact, and we speculate that these two residues may be involved in interdomain interactions in the full-length NC1 domain.

Collagen XVIII is a member of the so-called multiplexin family of collagens that also includes collagen XV. These two collagens are distinguished by C-terminal globular domains that share 58% sequence identity in their last

Table II. Refinement statistics

Resolution range (Å)	8–1.5
Reflections (working set)[a]	23521
Reflections (test set)[a]	2634
Non-H atoms	1416
R_{cryst}[b]	0.190
R_{free}[b]	0.224
R.m.s.d. bond lengths (Å)	0.009
R.m.s.d. bond angles (°)	1.6
R.m.s.d. B-factors for bonded atoms (Å^2)	3.7

[a]The test set is a randomly selected subset of the data (10%) that was not used in the refinement of the model. The working set comprises the remaining reflections.
[b]$R = \Sigma_h \|F_{obs}(h)\| - |F_{calc}(h)\|/\Sigma_h F_{obs}(h)$, where F_{obs} and F_{calc} are the observed and calculated structure factor amplitudes, respectively. R_{cryst} and R_{free} were calculated using the working set and the test set, respectively.
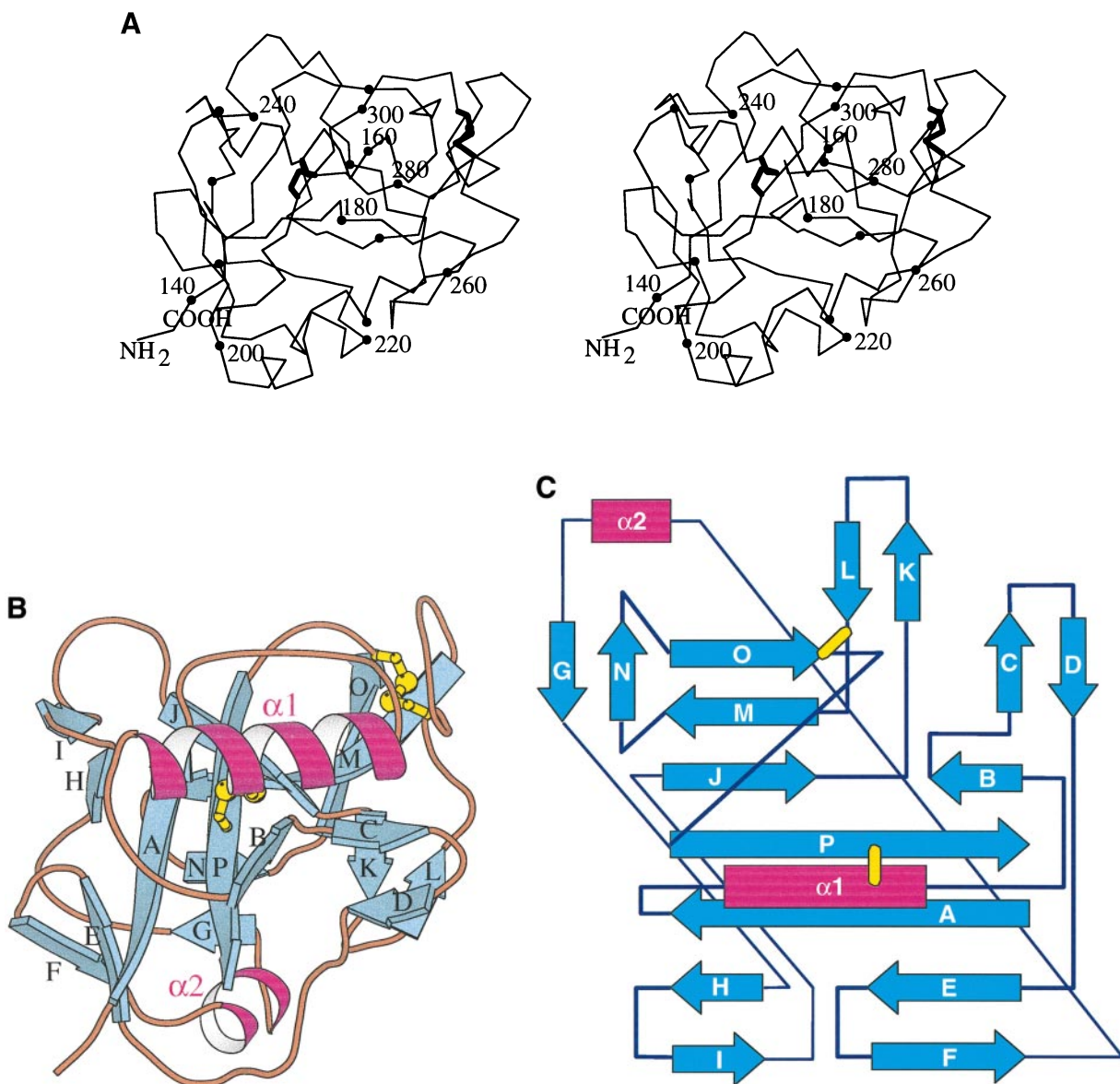
184 residues corresponding to endostatin in collagen XVIII (Oh *et al.*, 1994a). We suggest that this common region represents a new extracellular module (Bork *et al.*, 1996), as it appears unlikely that the compact endostatin structure is a soluble fragment excised from a larger domain. Interestingly, the ordered portion of the endostatin structure starts at an intron–exon boundary of α1(XVIII) collagen (Rehn *et al.*, 1996). The autonomous folding of endostatin is also indicated by the high production rate of recombinant mouse endostatin (10–20 mg/l per day).

Comparison with other structures

An automated search of the DALI database (Holm and Sander, 1993, 1994) unexpectedly revealed that endostatin resembles the carbohydrate recognition domain (CRD) of mammalian C-type lectins (Weis *et al.*, 1991; Drickamer, 1993). The highest scores (Z = 3.1) were obtained
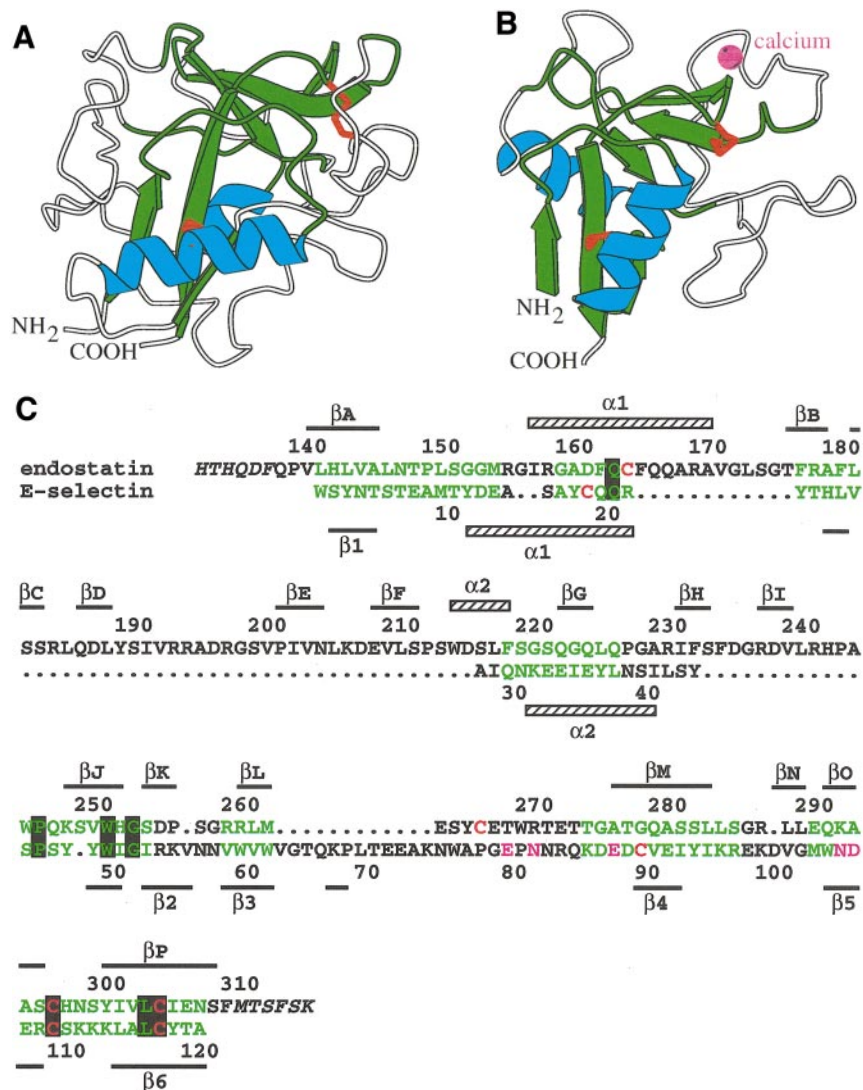

Fig. 3. (**A**) Stereo Cα-trace (Kraulis, 1991) of the endostatin structure. Spheres denote every tenth Cα-atom, and every 20th residue is labelled starting at Val140. Disulphide bridges are shown with thick bonds. (**B**) Cartoon representation (Kraulis, 1991) of the endostatin structure. The β-strands are sequentially labelled A–P. α-Helices are in pink and disulphide bridges are in yellow. (**C**) Topology diagram of the endostatin fold. The colour scheme for β-strands, α-helices and disulphide bridges is the same as in (B). Heavy and thin diagonal lines indicate connections above and below the large β-sheet, respectively.

Fig. 4. Cartoon representations (Kraulis, 1991) of (**A**) endostatin and (**B**) E-selectin (Graves *et al.*, 1994), emphasizing the structural similarities between endostatin and the C-type lectin CRD. The structurally analogous β-strand and loop regions equivalenced by DALI (Holm and Sander, 1993, 1994) are in green; disulphide bridges are in red and the calcium ion bound to E-selectin is shown as a pink sphere. (**C**) Structure-based alignment of the endostatin and E-selectin sequences. Structurally aligned residues are in green, cysteines are in red, and residues coordinating to the calcium ion in E-selectin are in pink. Endostatin residues in italics are not defined by the electron density. Amino acid identities are highlighted by black boxes. The secondary structure elements in the two structures are indicated above and below the alignment; the secondary structure definition for E-selectin is taken from Graves *et al.* (1994).

with E-selectin (Graves *et al.*, 1994) and lithostathine, a homologue of the CRD that does not bind calcium or carbohydrate ligands (Bertrand *et al.*, 1996). For the endostatin/E-selectin CRD pair, DALI identified 77 equivalent residues whose Cα-atoms could be superimposed with an r.m.s.d. of 3.1 Å. The corresponding sequence identity of 9% is well below the threshold of statistical significance, but the high degree of structural similarity strongly argues for an evolutionary relationship. Figure 4 shows that the entire β-sheet structure of the E-selectin CRD is contained within the endostatin structure; only strands E and F of the central β-sheet of endostatin do not have counterparts in the E-selectin CRD. The α-helices are in topologically equivalent positions in the two proteins, but their disposition relative to the central β-sheet varies. Significantly, both disulphide bridges in endostatin and E-selectin align almost perfectly when the

three-dimensional structures are superimposed. Apart from Cys296 and Cys304, there are five additional amino acid residues in endostatin that have identical counterparts in the E-selectin structure. With the exception of the surface residue Gln163 whose conservation may be adventitious, the location of these identities is telling: Pro246 and Gly253 flank the crucial strand J in the large β-sheet, and Trp251 forms the nucleus of the smaller of the two hydrophobic cores in endostatin. The equivalent region in the C-type lectin CRD domain is the 'WIGL' sequence (aromatic-aliphatic-glycine-aliphatic), which is an important component of the C-type lectin consensus (Weis *et al.*, 1991; Drickamer, 1993).

The differences between endostatin and the E-selectin CRD are concentrated in two regions, situated on opposite faces of the central β-sheet. In E-selectin, the connection between α1 and strand β2 (the equivalent of endostatin
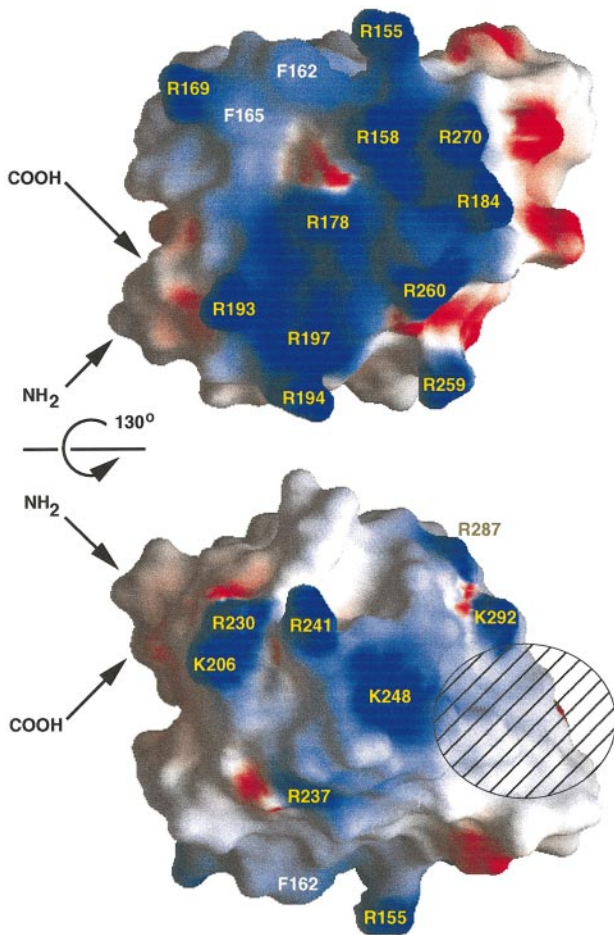
Fig. 5. Electrostatic surface representation (Nicholls, 1992) of endostatin. The two views of the endostatin surface are related by a rotation of 130° about the horizontal axis. Blue indicates regions of positive potential and red regions of negative potential, at the 5 kT/e level. The N- and C-termini are indicated; basic residues and the solvent-exposed side chains of Phe162 and Phe165 are labelled in yellow and white, respectively. The hatched area corresponds to the oligosaccharide-binding site of C-type lectin CRDs.

βJ) is afforded by a short strand, helix α2 and the loops preceding and following α2. The equivalent, much more elaborate region in endostatin contains eight, predominantly short strands (B–I) as well as α2 and accounts for most of the extra residues of endostatin compared with E-selectin. In this region there is little similarity between the two structures, apart from the general location of α2. In E-selectin, β5 (the equivalent of endostatin βO) and the loop preceding β4 form a high-affinity calcium-binding site involved in oligosaccharide binding (Graves *et al.*, 1994). There is no indication of calcium binding to endostatin and, indeed, we find that the long loop providing three of the five calcium ligands in E-selectin is shorter in endostatin and is arranged very differently, folded away from βO.

In summary, while endostatin is doubtless related to the C-type lectin CRD, it has lost one of the defining features of this protein family, namely calcium-dependent oligo-saccharide binding. This is not an unprecedented observation. We have already mentioned lithostathine, a CRD homologue that acts as an inhibitor of stone formation in the pancreas (Bertrand *et al.*, 1996). More relevantly, the

recently elucidated structure of the Link module showed that this hyaluronan-binding domain resembles the C-type lectin CRD but does not use calcium for glycosamino-glycan (GAG) binding (Kohda *et al.*, 1996). This is further discussed below in conjunction with the ligand-binding properties of endostatin.

A putative heparin-binding site

The mechanism(s) by which endostatin inhibits endothelial cell proliferation and angiogenesis have yet to be defined. However, given the high affinity of endostatin for heparin, interference with the heparan sulphate requirement of bFGF signalling is a distinct possibility. Endostatin contains a large number of basic residues, in particular arginines, and their distribution on the protein surface provides a clue to the location of the heparin-binding site. We note that of the 15 arginine residues present in mouse endostatin, all but one are fully conserved in the human protein, the exception being a conservative replacement by lysine (the overall identity between mouse and human endostatin is 87%; Oh *et al.*, 1994b). Given the general location of arginines in surface loops, this high degree of conservation is noteworthy and hints at an important function.

Apart from two crystal structures of bFGF with bound heparin-derived tetra- and hexasaccharides (Faham *et al.*, 1996), nothing is known about the detailed structural requirements of heparin binding to proteins. However, it is clear that basic residues are crucial; these are often arranged as discrete clusters with a spacing that matches the distribution of GAG sulphate groups (Fromm *et al.*, 1997). In addition, neutral polar residues are required to provide hydrogen bonding partners for the sugar moieties (Thompson *et al.*, 1994). The importance of aromatic residues, which is well established in the recognition of uncharged oligosaccharides, is less evident in the case of GAGs.

A representation of the surface electrostatic properties of endostatin is shown in Figure 5. Eleven out of the total of 15 arginine residues cluster on one face of the molecule (Figure 5). This extensive basic patch (diameter ⊤20 Å) involves α1 and α2, strand B, the long loop connecting the C–D β-hairpin to strand E, as well as strand L and the following loop around Cys266. The solvent-exposed side chains of Phe162 and Phe165 (see above) are found at the periphery of the patch. A detailed inspection highlights one particular area as a candidate heparin-binding site. The two arginine pairs Arg193/Arg194 and Arg259/Arg260 form the borders of a shallow depression, at the centre of which the side chain of Tyr215 emerges, surrounded by several additional polar residues. The residues forming this putative heparin-binding site are contributed by the long irregular loop preceding strand E and strand L, both elements unique to endostatin when compared with the C-type lectin CRD and the Link module. Binding of a larger GAG chain may involve additional arginines, possibly those centred around Arg158.

Endostatin contains a second, less extensive basic patch approximately opposite the large area defined by the 11 arginines discussed above (Figure 5). This patch, composed mainly of residues contributed by the H–I β-hairpin and the N-terminus of strand O, is of interest because it

is close to the ligand-binding site in the related C-type lectins CRD (Weis *et al.*, 1992; Graves *et al.*, 1994). We cannot fully dismiss the possibility that heparin may also bind to this region, although we note that some of the basic residues clearly serve structural purposes and would not be available for ligand binding (Arg230, Arg237 and Lys248). Furthermore, our assignment of the heparin-binding site to the face bearing exclusively arginine residues is consistent with results from chemical modification, which demonstrate the involvement in heparin binding of several arginine but no lysine residues (unpublished data).

The Link module, a domain of ~100 residues found in several extracellular matrix proteins and in the cell surface receptor CD44, is a distant relative of the C-type lectin CRD that binds the GAG hyaluronan. In the TSG-6 Link module structure (Kohda *et al.*, 1996), a critical basic residue in the loop preceding α1 and a patch of solvent-exposed aromatic side chains define a putative hyaluronan-binding site, which partly overlaps with the surface area used by the CRDs for oligosaccharide binding. In endostatin the equivalent region involves mainly strands M and O and the connecting loop to strand P, which do not coincide with either of the two basic patches described above. We therefore believe that endostatin and the Link module employ spatially distinct regions for GAG binding.

The function of the domain corresponding to endostatin in tissue-deposited collagen XVIII and XV is not known. Collagen XVIII is mainly found in vascular basement membrane regions (Muragaki *et al.*, 1995), and the C-terminal NC1 domain may mediate interactions with basement membrane GAGs or proteoglycans. Full-length collagen XVIII is likely to be immobilized in some kind of network. Proteolytic cleavage in the NC1 domain, perhaps by a proteinase secreted by a tumour, could produce soluble endostatin which would be free to diffuse to its targets and elicit its effects on endothelial cell proliferation and angiogenesis. The intact NC1 domain of collagen XVIII is not an inhibitor of angiogenesis (unpublished data), corroborating earlier indications that the antiproliferative activity of endostatin may be cryptic (O'Reilly *et al.*, 1997). It is not obvious from our structure how this regulation of activity may take place. Steric blocking of an important epitope by the N-terminal portion of the NC1 domain is a possibility. However, other scenarios can be envisaged. For instance, the exposed and mobile polypeptide chain termini of endostatin may be important for the antiproliferative effect.

Guided by our structure we are now planning to delineate the heparin-binding site of endostatin by site-directed mutagenesis, and this will provide a powerful tool to define and dissect the mechanisms of action of endostatin. If endostatin indeed acts by interfering with the heparan sulphate requirement of bFGF signalling, abolishment of heparin binding would be expected to be accompanied by a loss of inhibition of endothelial cell proliferation and/or angiogenesis. Alternatively, endostatin may interact with an as yet undefined protein or receptor and cause inhibition independent of GAG binding. Heparin binding may also turn out to be only one of several critical components of the mechanism, as is the case with bFGF signalling. Finally, the question how endostatin activity is regulated by the proteolytic unmasking of cryptic epitopes

will be addressed by the functional characterization and structure determination of the full-length NC1 domain of collagen XVIII. The insights gained from these studies should prove useful for the development of new approaches to cancer therapy.

Materials and methods

Construction of expression vector
Mouse α1(XVIII) cDNA clone mc3b (Oh *et al.*, 1994a) was used as a template to amplify the sequence encoding endostatin by polymerase chain reaction (PCR) with Vent polymerase (New England Biolabs) following the manufacturer's instructions. The primer for the 5′ end was GTCAGCTAGCTCATACTCATCAGGAC and that for the 3′ end was GTCACTCGAGCTATTTGGAGAAAGAGGTC. In addition to the annealing sequences the primers contained an *Nhe*I site at the 5′ end or a stop codon followed by an *Xho*I site at the 3′ end, in order to allow the in-frame insertion of the construct into the BM-40 signal peptide (Mayer *et al.*, 1993). The PCR fragment was cloned into the modified episomal expression vector pCEP-Pu (Kohfeldt *et al.*, 1997). The sequence of the construct was confirmed by cycle sequencing using Dye Terminator Cycle Sequencing Ready Reaction Kit (ABI).

Expression and purification of recombinant mouse endostatin
Human embryonic kidney cells that express the EBNA-1 protein from Epstein–Barr virus (293-EBNA cells, Invitrogen) were used for transfection with the expression vector (Kohfeldt *et al.*, 1997). Resistant cells were selected with puromycin (0.5 μg/ml) and used for collection of serum-free conditioned medium. The medium (⊤1 l) was dialysed against 0.1 M NaCl, 0.05 M Tris–HCl pH 7.4, and then applied onto a heparin–Sepharose CL-6B column (2.5×20 cm, Pharmacia) equilibrated in the same buffer. A linear 0.1–1.0 M NaCl gradient (500 ml) was used for elution. Endostatin eluted at 0.4–0.5 M NaCl and was further purified on a Superose 12 column (HR16/50, Pharmacia) equilibrated in 0.2 M ammonium acetate, pH 6.8. The purified product was soluble in neutral buffer and showed a single 22 kDa band in SDS gel electrophoresis under reducing conditions. The protein has a single N-terminal sequence APLAHTHQ and contains less than one residue of hexosamine per molecule.

Crystallization and data collection
Crystals were obtained at room temperature by the hanging drop vapour diffusion method. Equal volumes (typically 2 μl) of a 10 mg/ml solution of endostatin in 5 mM MOPS, pH 6.8, and 1.5–1.7 M ammonium phosphate, pH 4.7–5.3, were mixed and equilibrated against 1 ml of the latter solution. The crystals belong to space group $P2_12_12_1$ with unit cell constants $a = 45.6$ Å, $b = 54.0$ Å, $c = 65.9$ Å. There is one molecule of endostatin in the asymmetric unit, resulting in a solvent content of 37%. For heavy-atom soaks, crystals were stabilized in 1.8 M Li_2SO_4, 0.1 M Na-acetate, pH 5.3. All diffraction data except native II were collected at room temperature using an MAR image plate detector mounted on a rotating anode generator operated at 4 kW (CuKα radiation, $\lambda = 1.54$ Å). For derivative data collection, crystals were rotated around their carefully aligned *a* axis to minimize systematic errors in the measurement of Bijvoet pairs. Native II data were collected at room temperature on beamline 9.6 of the Daresbury Synchrotron Radiation Source using an MAR image plate detector ($\lambda = 0.87$ Å). Data were integrated with MOSFLM (Leslie, 1994) and reduced with programs of the CCP4 suite (Collaborative Computing Project No. 4, 1994). Data collection statistics are summarized in Table I.

Structure solution and refinement
Three heavy-atom derivatives and the native I data were used for phasing by the MIRAS method (Table I). Soak conditions were 3 mM UO_2SO_4 for 3 days, 20 mM $K_2Pt(CN)_4$ for 1 day and 10 mM $NaAu(CN)_2$ for 2 days. Heavy-atom sites were deduced from difference Patterson maps, brought to a common origin and hand by cross-phased difference Fourier maps, and refined with MLPHARE (Z.Otwinowski; Collaborative Computing project No. 4, 1994). Due to the high isomorphism of the U and Pt derivatives, useful MIRAS phases could be obtained to a resolution of 2.2 Å (mean figure-of-merit 0.584). The MIRAS map was subjected to density modification with DM (Cowtan and Main, 1996) in 'combine omit' mode employing solvent flattening, histogram matching and Sayre's equation. Approximately 75% of the structure could be built with

confidence into the resulting map using O (Jones *et al.*, 1991). The remaining loop structures were added after combination of partial model phases with the experimental phases using SIGMAA (Read, 1986). The structure was first refined with X-PLOR (Brünger, 1992) against the native I data at 2.0 Å resolution to $R_{cryst} = 0.192$ ($R_{free} = 0.237$). Refinement against the synchrotron native II data was then initiated by a round of simulated annealing refinement starting from 3000 K to remove model bias, followed by conventional positional and *B*-factor refinement. The final model comprises residues 138–309 and 83 water molecules (Table II); 86.3% of the amino acid residues are in the most favourable regions of the Ramachandran plot, with the remaining 13.7% in additionally allowed regions, as defined by PROCHECK (Laskowski *et al.*, 1993).

Coordinates and structure factors have been deposited in the Brookhaven Protein Data Bank (accession code 1KOE) and will be held for one year after publication.

Acknowledgements

References

Beck,L.,Jr and D'Amore,P.A. (1997) Vascular development: cellular and molecular regulation. *FASEB J.*, **11**, 365–373.

Bertrand,J.A., Pignol,D., Bernard,J.-P., Verdier,J.-M., Dagorn,J.-C. and Fontecilla-Camps,J.C. (1996) Crystal structure of human lithostathine, the pancreatic inhibitor of stone formation. *EMBO J.*, **15**, 2678–2684.

Boehm,T., Folkman,J., Browder,T. and O'Reilly,M.S. (1997) Anti-angiogenic therapy of experimental cancer does not induce acquired drug resistance. *Nature*, **390**, 404–407.

Bork,P., Downing,A.K., Kieffer,B. and Campbell,I.D. (1996) Structure and distribution of modules in extracellular proteins. *Q. Rev. Biophys.*, **29**, 119–167.

Brown,K.J. and Parish,C.P. (1994) Histidine-rich glycoprotein and platelet factor 4 mask heparan sulfate proteoglycans recognized by acidic and basic fibroblast growth factor. *Biochemistry*, **33**, 13918–13927.

Brünger,A.T. (1992) *X-PLOR version 3.1: A System for Crystallography and NMR*. Yale University Press, New Haven, CT.

Bussolino,F., Mantovani,A. and Persico,G. (1997) Molecular mechanisms of blood vessel formation. *Trends Biochem. Sci.*, **22**, 251–256.

Clapp,C., Martial,J.A., Guzman,R.C., Rentier-Delrue,F. and Weiner,R.I. (1993) The 16 kDa N-terminal fragment of human prolactin is a potent inhibitor of angiogenesis. *Endocrinology*, **133**, 1292–1299.

Collaborative Computing Project No. 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr.*, **D50**, 760–763.

Cowtan,K.D. and Main,P. (1996) Phase combination and cross validation in iterated density-modification calculations. *Acta Crystallogr.*, **D52**, 43–48.

Dameron,K.M., Volpert,O.V., Tainsky,M.A. and Bouck,N. (1994) Control of angiogenesis in fibroblasts by p53 regulation of thrombospondin-1. *Science*, **265**, 1582–1584.

Drickamer,K. (1993) Ca^{2+}-dependent carbohydrate-recognition domains in animal proteins. *Curr. Opin. Struct. Biol.*, **3**, 393–400.

Faham,S., Hileman,R.E., Fromm,J.R., Linhardt,R.J. and Rees,D.C. (1996) Heparin structure and interactions with basic fibroblast growth factor. *Science*, **271**, 1116–1120.

Ferrara,N., Clapp,C. and Weiner,R.I. (1991) The 16 kDa fragment of prolactin specifically inhibits basal or fibroblast growth factor-stimulated growth of capillary endothelial cells. *Endocrinology*, **129**, 896–900.

Friesel,R.E. and Maciaq,T. (1995) Molecular mechanisms of angiogenesis: fibroblast growth factor signal. *FASEB J.*, **9**, 919–925.

Fromm,J.R., Hileman,R.E., Caldwell,E.E.O., Weiler,J.M. and Linhardt, R.J. (1997) Pattern and spacing of basic amino acids in heparin binding sites. *Arch. Biochem. Biophys.*, **343**, 92–100.

Good,D.J., Polversini,P.J., Rastinejad,F., Le Beau,M.M., Lemons,R.S., Frazier,W.A. and Bouck,N.P. (1990) A tumor suppressor-dependent inhibitor of angiogenesis is immunologically and functionally indistinguishable from a fragment of thrombospondin. *Proc. Natl Acad. Sci. USA*, **87**, 6624–6628.

Graves,B.J. *et al.* (1994) Insight into E-selectin/ligand interaction from the crystal structure and mutagenesis of the lec/EGF domains. *Nature*, **367**, 532–538.

Hanahan,D. and Folkman,J. (1996) Patterns and emerging mechanisms of the angiogenic switch during tumorigenesis. *Cell*, **86**, 353–364.

Herr,A.B., Ornitz,D.M., Sasisekharan,R., Venkataraman,G. and Waksman,G. (1997) Heparin-induced self-association of fibroblast growth factor-2. Evidence for two oligomerization processes. *J. Biol. Chem.*, **272**, 16382–16389.

Holm,L. and Sander,C. (1993) Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.*, **233**, 123–138.

Holm,L. and Sander,C. (1994) The FSSP database of structurally aligned protein fold families. *Nucleic Acids Res.*, **22**, 3600–3609.

Homandberg,G.A., Williams,J.E., Grant,D., Schuhmacher,B. and Eisenstein,R. (1985) Heparin-binding fragments of fibronectin are potent inhibitors of endothelial cell growth. *Am. J. Pathol.*, **120**, 327–332.

Jones,T.A., Zou,J.-Y., Cowan,S.W. and Kjeldgaard,M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr.*, **A47**, 110–119.

Kohda,D., Morton,C.J., Parkar,A.A., Hatanaka,H., Inagaki,F.M., Campbell,I.D. and Day,A.J. (1996) Solution structure of the Link module: a hyaluronan-binding domain involved in extracellular matrix stability and cell migration. *Cell*, **86**, 767–775.

Kohfeldt,E., Maurer,P., Vannahme,C. and Timpl,R. (1997) Properties of the extracellular calcium binding module of the proteoglycan testican. *FEBS Lett.*, **414**, 557–561.

Kraulis,P.J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.*, **24**, 946–950.

Laskowski,R.A., MacArthur,M.W., Moss,D.S. and Thornton,J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, **26**, 283–291.

Leslie,A.G.W. (1994) *MOSFLM Users Guide*. MRC-LMB, Cambridge, UK.

Mayer,U., Nischt,R., Pöschl,E., Mann,K., Fukuda,K., Gerl,M., Yamada,Y. and Timpl,R. (1993) A single EGF-like motif of laminin is responsible for high-affinity nidogen binding. *EMBO J.*, **12**, 1879–1885.

Moy,F.J., Safran,M., Seddon,A.P., Kitchen,D., Böhlen,P., Aviezer,D., Yayon,A. and Powers,R. (1997) Properly oriented heparin-decasaccharide-induced dimers are the biologically active form of basic fibroblast growth factor. *Biochemistry*, **36**, 4782–4791.

Muragaki,Y., Timmons,S., Griffith,C.M., Oh,S.P., Fadel,B., Quertermous,T. and Olsen,B.R. (1995) Mouse *Col18a1* is expressed in a tissue-specific manner as three alternative variants and is localized in basement membrane zones. *Proc. Natl Acad. Sci. USA*, **92**, 8763–8767.

Myers,J.C., Kivirikko,S., Gordon,M.K., Dion,A.S. and Pihlajaniemi,T. (1992) Identification of a previously unknown human collagen chain, α1(XV), characterized by extensive interruptions in the triple-helical region. *Proc. Natl Acad. Sci. USA*, **89**, 10144–10148.

Nicholls,A. (1992) *GRASP: Graphical Representation and Analysis of Surface Properties*. Columbia University, NY.

Oh,S.P., Kamagata,Y., Muragaki,Y., Timmons,S., Ooshima,A. and Olsen,B.R. (1994a) Isolation and sequencing of cDNAs for proteins with multiple domains of Gly-Xaa-Yaa repeats identify a distinct family of collageneous proteins. *Proc. Natl Acad. Sci. USA*, **91**, 4229–4233.

Oh,S.P., Warman,M.L., Seldin,M.F., Cheng,S.-D., Knoll,J.H.M., Timmons,S. and Olsen,B.R. (1994b) Cloning of cDNA and genomic DNA encoding human type XVIII collagen and localization of the α1(XVIII) collagen gene to mouse chromosome 10 and human chromosome 21. *Genomics*, **19**, 494–499.

O'Reilly,M.S. *et al.* (1994) Angiostatin: a novel angiogenesis inhibitor that mediates the suppression of metastases by a Lewis lung carcinoma. *Cell*, **79**, 315–328.

O'Reilly,M.S. *et al.* (1997) Endostatin: an endogenous inhibitor of angiogenesis and tumor growth. *Cell*, **88**, 277–285.

Read,R.J. (1986) Improved Fourier coefficients for maps using phases from partial structure factors with errors. *Acta Crystallogr.*, **A42**, 140–149.

Rehn,M. and Pihlajaniemi,T. (1994) α1(XVIII), a collagen chain with frequent interruptions in the collageneous sequence, a distinct distribution and homology with type XV collagen. *Proc. Natl Acad. Sci. USA*, **91**, 4234–4238.

Rehn,M. and Pihlajaniemi,T. (1995) Identification of three N-terminal ends of type XVIII collagen chains and tissue-specific differences in the expression of the corresponding transcripts. *J. Biol. Chem.*, **270**, 4705–4711.

Rehn,M., Hintikka,E. and Pihlajaniemi,T. (1996) Characterization of the mouse gene for the α1 chain of type XVIII collagen (*Col18a1*) reveals that the three variant N-terminal polypeptide forms are transcribed from two widely separated promoters. *Genomics*, **32**, 436–446.

Risau,W. (1997) Mechanisms of angiogenesis. *Nature*, **386**, 671–674.

Sage,E.H. (1997) Pieces of eight: bioactive fragments of extracellular proteins as regulators of angiogenesis. *Trends Cell Biol.*, **7**, 182–186.

Thompson,L.D., Pantoliano,M.W. and Springer,B.A. (1994) Energetic characterization of the basic fibroblast growth factor-heparin interaction: identification of the heparin-binding domain. *Biochemistry*, **33**, 3831–3840.

Tolsma,S.S., Volpert,O.V., Good,D.J., Frazier,W.A., Polverini,P.J. and Bouck,N. (1993) Peptides derived from two separate domains of the matrix protein thrombospondin-1 have antiangiogenic activity. *J. Cell Biol.*, **122**, 497–511.

Weis,W.I., Kahn,R., Fourme,R., Drickamer,K. and Hendrickson,W.A. (1991) Structure of the calcium-dependent lectin domain from a rat mannose-binding protein determined by MAD phasing. *Science*, **254**, 1608–1615.

Weis,W.I., Drickamer,K. and Hendrickson,W.A. (1992) Structure of a C-type mannose-binding protein complexed with an oligosaccharide. *Nature*, **360**, 127–134.

Yayon,A., Klagsbrun,M., Esko,J.D., Leder,P. and Ornitz,D.M. (1991) Cell surface, heparin-like molecules are required for binding of basic fibroblast growth factor to its high affinity receptor. *Cell*, **64**, 841–848.