The EMBO Journal vol.15 no.24 pp.6789–6797, 1996

Crystal structure of coagulogen, the clotting protein from horseshoe crab: a structural homologue of nerve growth factor

Andreas Bergner[1,2], Vaheh Oganessyan[1], Tatsushi Muta[3], Sadaaki Iwanaga[3], Dieter Typke[4], Robert Huber[1] and Wolfram Bode[1]

[1]Max-Planck-Institut für Biochemie, Abteilung Strukturforschung and [4]Max-Planck-Institut für Biochemie, Abteilung Molekulare Strukturbiologie, Am Klopferspitz 18a, D-82152 Martinsried, Germany, and [3]Department of Biology, Faculty of Science, Kyushu University 33, Fukuoka 812-81, Japan

[2]Corresponding author

The clotting cascade system of the horseshoe crab (*Limulus*) is involved in both haemostasis and host defence. The cascade results in the conversion of coagulogen, a soluble protein, into an insoluble coagulin gel. The clotting enzyme excises the fragment peptide C from coagulogen, giving rise to aggregation of the monomers. The crystal structure of coagulogen reveals an elongated molecule that embraces the helical peptide C fragment. Cleavage and removal of the peptide C would expose an extended hydrophobic cove, which could interact with the hydrophobic edge of a second molecule, leading to a polymeric fibre. The C-terminal half of the coagulogen molecule exhibits a striking topological similarity to the neurotrophin nerve growth factor (NGF), providing the first evidence for a neurotrophin fold in invertebrates. Similarities between coagulogen and Spätzle, the *Drosophila* ligand of the receptor Toll, suggest that the neurotrophin fold might be considered more ancient and widespread than previously realized.

Keywords: crystal structure/cystine knot/horseshoe crab/nerve growth factor/Spätzle

Introduction

The survival of a multicellular organism depends upon effective defence systems for recognition and elimination of foreign microorganisms. In addition to cytolysis, cell agglutination, phagocytosis and antimicrobial activities, horseshoe crabs possess a haemolymph coagulation system, which participates both in defence against microbes and in haemostasis (Muta and Iwanaga, 1996). This coagulation system is composed of five protein components which make up a clotting cascade system (Iwanaga, 1993b), the serine proteinase zymogens factors C, B, G, the proclotting enzyme and the clottable protein coagulogen. These clotting factors are sequestered in the cytoplasmic L-granules and are released into haemolymph through rapid exocytosis upon stimulation. The clotting factors C (Muta *et al.*, 1993b) and G (Muta *et al.*, 1995) act as highly sensitive biosensors for lipopolysaccharides (LPS) and (1,3)-β-glucan, present on the outer membrane of Gram-negative bacteria and the cell walls of fungi, respectively. Invading pathogens trigger these biosensors, resulting in the sequential activation of the clotting factors and subsequent clot formation. The extreme sensitivity of the clotting cascade for LPS is used in the so-called 'limulus test', a widely employed assay method for bacterial endotoxins or LPS (Levin and Bang, 1964; Tanaka and Iwanaga, 1993).

In the final step of the cascade activation, the clotting enzyme converts coagulogen, a 175 amino acid single chain polypeptide (19.6 kDa), to insoluble coagulin by limited proteolysis at two sites (Arg18–Thr19 and Arg46–Gly47). Excision of the intermediate peptide C (Thr19–Arg46) results in the coagulin monomer AB, consisting of the N-terminal fragment A (Ala1–Arg18) and the C-terminal fragment B (Gly47–Phe175) covalently linked via two disulfide bridges. Coagulin is capable of self-aggregating to a gel-like substance composed of polymerized AB monomers. The coagulin gel hinders the spread of invaders by immobilizing them, contributing to the host defence system of the horseshoe crab in addition to sealing haemolymph leakages upon injury. Antimicrobial substances such as tachyplesins, anti-LPS factor and big defensin are also released from the haemocytes, liquidating the engulfed pathogens. These functions of the *Limulus* clotting cascade point to a common evolutionary origin of the immune defence and blood clotting systems. By analogy to the mammalian clotting system (Iwanaga, 1993a), the clotting enzyme corresponds to thrombin, and coagulogen to fibrinogen. This is only a formal analogy, however, as fibrinogen and coagulogen are proteins of quite different size and amino acid sequence. Moreover, factor XIIIa covalently cross-links fibrinogen monomers, whereas polymerized coagulin contains no covalent cross-links. The X-ray structure of coagulogen from Japanese horseshoe crab (*Tachypleus tridentatus*) described herein provides a structural basis for understanding the polymerization mechanism. Surprisingly, we found that the C-terminal domain of coagulogen exhibits a nerve growth factor (NGF)-like fold, suggesting new and unexpected aspects of the evolution of protein folds.

Results

Overall structure

The crystals contain two coagulogen molecules per asymmetric unit, with the monomers related by a local 2-fold screw axis. Ultracentrifugal sedimentation equilibrium analysis reveals only monomeric coagulogen in solution (Nakamura *et al.*, 1976). The dimer contacts also exhibit no main chain interactions, indicating that the observed dimerization is a crystallization artefact. As the monomers are quite similar in structure, the description of the structure will be confined to one monomer.

© Oxford University Press

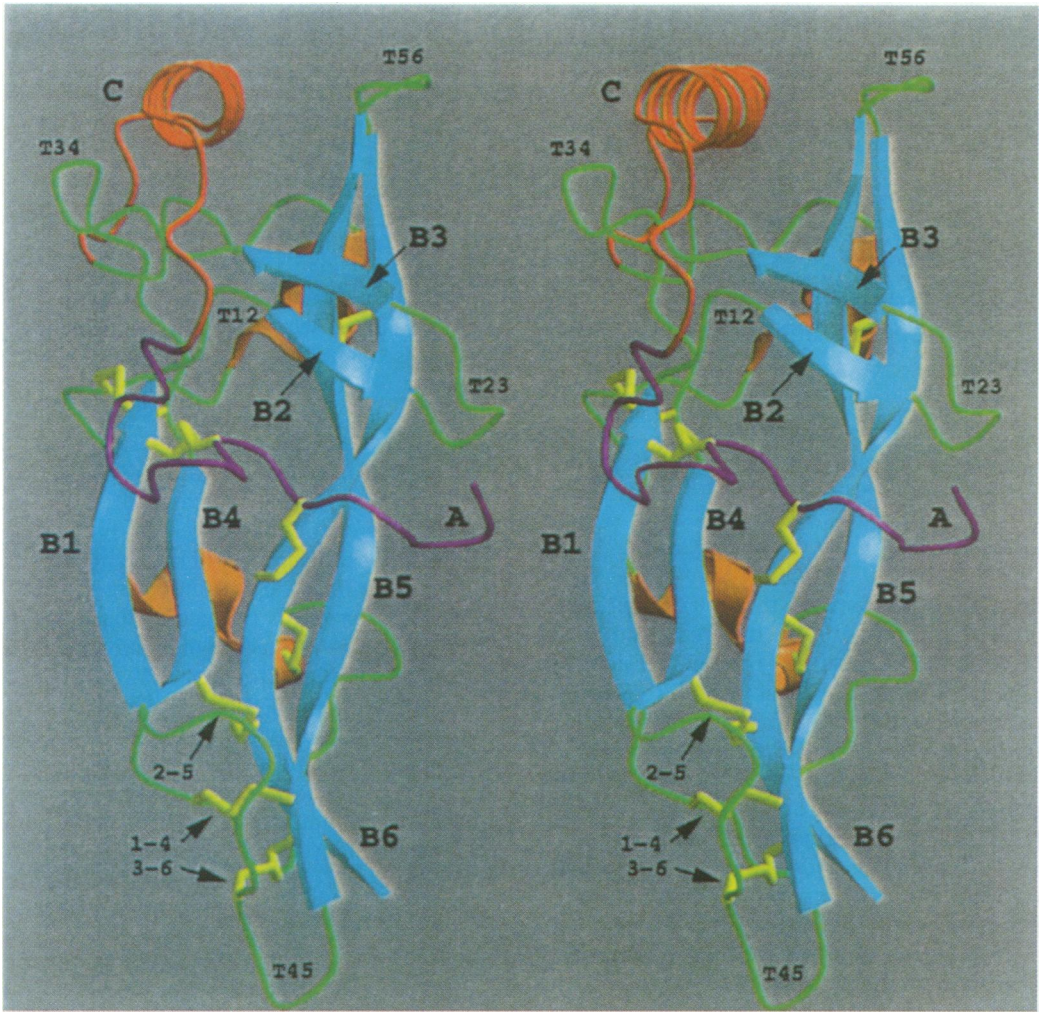

Fig. 1. Stereo view (Kraulis, 1991) of a coagulogen monomer, showing fragments A, C and B and the secondary structure: The structure is dominated by the β-strands (blue, labelled sequentially B1–B6) and multiple coils and turns (green, labelled T12, T23, T34, T45 and T56) of fragment B. Short helical segments (orange) in fragment B are in the background. The mainly α-helical peptide C (red), which is released upon cleavage, covers a reasonable part of the surface at the top. Its characteristic helix is embraced like a cylinder in a tong formed by segments of fragment B. The N-terminal fragment A (violet) is connected to fragment B by two disulfide bridges (yellow). The whole cysteine-rich structure possesses eight disulfide bridges. Six cysteines (1–6) form a characteristic arrangement called a cystine knot, whereby the disulfide bridge (1–4) penetrates a macrocyclic ring formed by the bridges (2–5), (3–6) and the polypeptide chain. Note that the cleavage sites themselves were not observed in the electron density, but are shown for convenience.

Coagulogen is an elongated molecule

Each coagulogen monomer is an elongated molecule of approximate dimensions 60, 30 and 20 Å. In the orientation depicted in Figure 1, the lower part represents fragments A and B, whereas the helix at the top belongs to the peptide C fragment. The shape of the molecule is reminiscent of a tong, with its two arms (fragment AB) grasping the helix between its tips. This four and a half turn α-helix comprises most of the peptide C fragment. Fragment AB consists mainly of β-structure, with some irregular multiple turn and coil elements. These secondary structures are in general agreement with a secondary structure prediction (Srimal *et al.*, 1985). The C-terminal portion of fragment B, designated BII (Cys88–Phe175), is made up of six β-strands, designated B1–B6, and large loop regions between them, designated T12, T23, T34, T45 and T56.

The two β-strands B6 (Gly158–Arg173) and B5 (Arg139–Asn153) and the loop between them (T56) shape the right 'arm' of fragment AB, forming an extremely long characteristic extended antiparallel β-sheet, β1, with an overall twist of 360°. The regular hydrogen bond pattern is interrupted by a β-bulge due to the insertion of an additional residue (Cys168) (see Figure 2). Two two-stranded antiparallel β-sheets, β2 and β3, form the second 'arm' on the left hand side. β2, consisting of the β-strands B4 (Val122–Arg128) and B1 (Val90–Val96), is oriented parallel to β1. β3, in contrast, built from the β-strands B2 (Phe99–Thr102) and B3 (Phe107–Val111), is oriented perpendicular to the long molecular axis crossing the β1 sheet in the foreground and fixed to the β1 sheet through two main chain hydrogen bonds.

The C-terminal BII segment forms a rigid skeleton surrounded by the N-terminal half of the molecule, consisting of fragments A and C and the N-terminal part of fragment B, BI (Gly47–Glu87). The BI fragment, which sits at the back of the BII skeleton, serves as a linking fragment between the rigid BII domain and the C-terminal cleavage site of peptide C. Fragment BI is anchored to

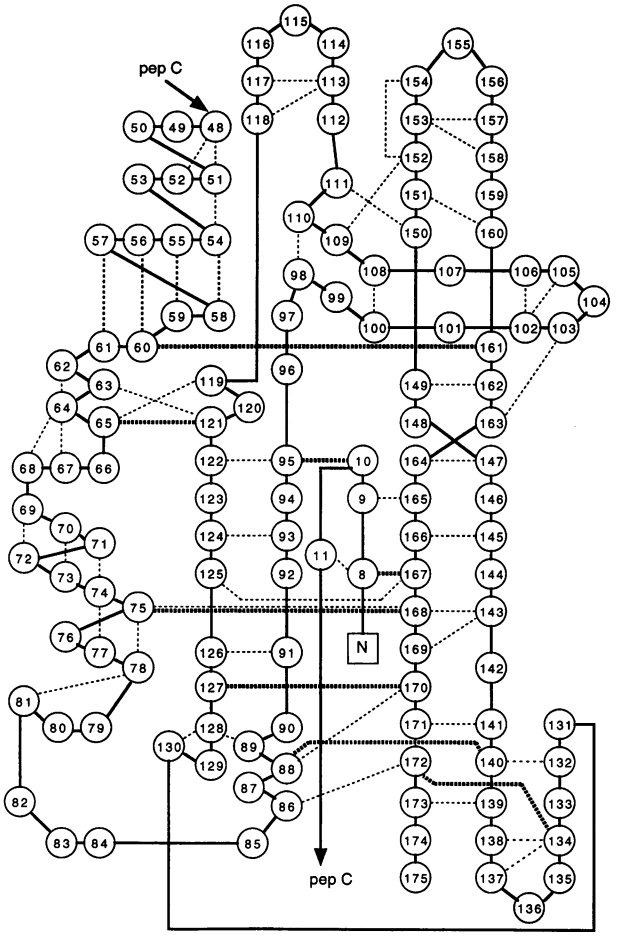

Fig. 2. Main chain interaction scheme for coagulogen: the connectivity pattern between all 16 half-cystines is eludicated. The six β-strands B1–B6 form three antiparallel β-sheets, β1 (B5 and B6), β2 (B1 and B4) and β3 (B2 and B3). β1 shows a large twist and bulges at residue number 168 (Cys). The eight disulfide bridges (Cys8–Cys167, Cys10–Cys95, Cys60–Cys161, Cys65–Cys121, Cys75–Cys168, Cys88–Cys140, Cys127–Cys170, Cys134–Cys172) are depicted by thick dashed lines. The bridges Cys88–Cys140 (1–4), Cys127–Cys170 (2–5) and Cys134–Cys172 (3–6) form the cystine knot. The dotted lines represent main chain hydrogen bonds. (Per pair a maximum of one hydrogen bond is displayed.) The part between residues 11 and 48 belongs either to the peptide C fragment, which is not connected to the rest of the molecule by main chain interactions, or is not defined in the electron density, and therefore is neglected in this scheme.

the BII domain by three disulfide bridges, and consists of multiple turn elements, and a short α-helical (Pro55–Gly61) and two 3_{10}-helical (Pro70–Ser73, Cys75–Asn77) segments.

The six N-terminal residues Ala1–Pro6 and the C-terminal segment of fragment A leading into peptide C are poorly if at all defined by electron density and are, therefore, disordered. The well-defined part of fragment A (Ile7–Glu12) runs along the molecular surface approximately perpendicular to the two large β-sheets β1 and β2 of the BII domain, to which it is connected by two disulfide bridges. At the top, the α-helix of peptide C (Thr24–Glu40) extends from the front to the back, covering a reasonable part of the surface provided by the AB fragment. The interface between peptide C and the AB fragment is ~680 Å^2 compared with the 8920 Å^2 surface of the whole coagulogen protomer. The N- and C-terminal extensions of the peptide C helix which connect it to both cleavage sites exhibit irregular structures. The peptide regions flanking each cleavage site (Pro13–Thr23, Gly42–Ser49) seem to be of a high mobility, with residues Gly17–Val22 and Gly42–Gly47 completely undefined by electron density. Presumably, these fragments can adapt easily to the extended substrate binding site of the active clotting enzyme, allowing rapid cleavage.

The two 'tips' of domain BII (loops T34 and T56) and the N-terminal portion of the BI part (around Phe48) together form a cove in the upper portion of the coagulogen molecule, which partially encloses the peptide C helix. The surface of this cove, lined by Phe48, Ile50, Phe51, Phe57 (BI), Val109, Val111, Phe118 (BII,T34) and Val150, Tyr152, Leu154, Phe159 (BII,T56), is quite hydrophobic in nature. Removal of peptide C would expose this hydrophobic surface to bulk solvent. In contrast, the peptide C helix exhibits amphiphilic properties, with exclusively hydrophobic residues (Ile30, Val34 and Val37) pointing to the hydrophobic cove of the AB fragment, and several charged hydrophilic residues (Asp28, Glu31, Lys32, Gln39 and Glu40) exposed to the bulk solvent. This hydrophobic interface, sandwiched between the peptide C helix and the hydrophobic layer of the cove, extends deep into the centre of the coagulogen molecule and forms a major hydrophobic core region in the upper part of the molecule. A second small core region is located in the lower part of the molecule, dominated by three characteristically arranged disulfide bridges that stabilize the molecule by connecting the β-sheets β1 and β2.

Self-aggregation of coagulogen

The exposure of the hydrophobic cove upon cleavage and release of the helical peptide C fragment suggests a mechanism of polymerization, involving other exposed hydrophobic residues of the AB fragment. Although no large accumulation of hydrophobic residues upon the molecular surface of coagulogen stands out, two hydrophobic surface patches may be involved in gel formation. The first one comprises Tyr76, Phe78, Phe81 and Phe84 of the BI segment, while the second one is made up of Phe99, Tyr101 and Phe107 of the BII domain. Other exposed hydrophobic residues such as Trp123 or Tyr136 appear as single hydrophobic islands. After release of fragment C, the hydrophobic cove of the AB monomer would be flanked by the side chains of Glu155 and Arg115, giving rise to a distinctive charge–hydrophobic–charge pattern at the surface (see Figure 3). As the clotting reaction appears to be pH dependent, electrostatic effects might also play a role in the aggregation of the AB monomers.

Electron microscopy of coagulin fibres

We have carried out a preliminary study of the structural organization of the coagulin gel by electron microscopy. It was found that the gel-like structure is made up of a network of linear fibres of varying thickness with crossings and branching points (see Figure 4a). Along the fibres, a periodicity of ~45 Å can be observed, verified by digitizing the images and calculating the Fourier transform (Figure 4b and c). The power spectrum shows a broad ring indicating that the 45 Å periodicity is poorly defined. In

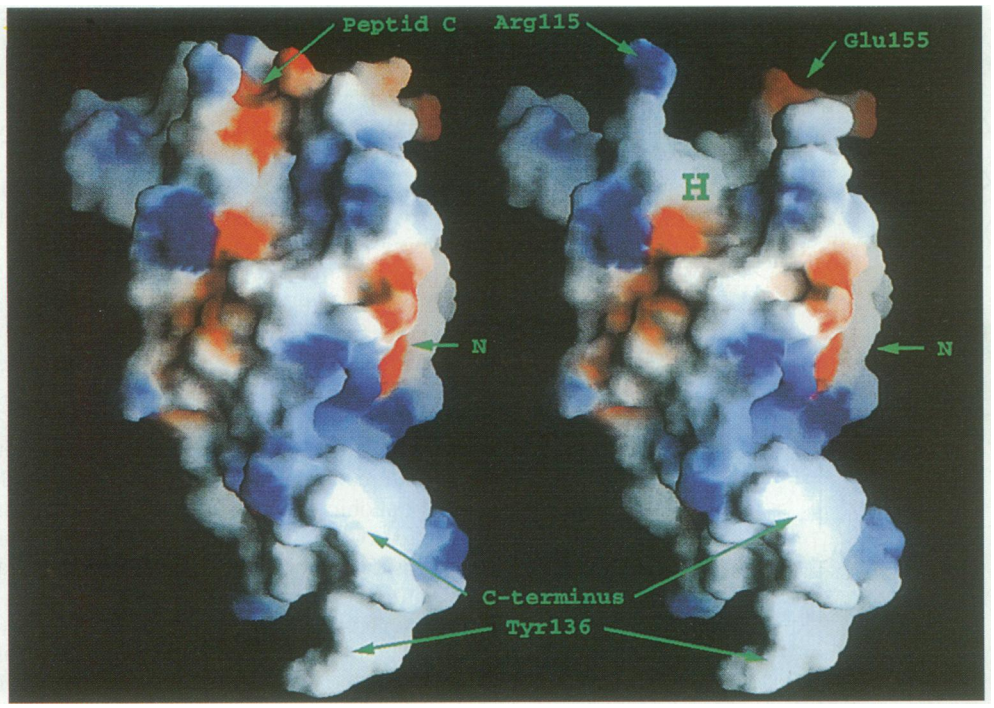

Fig. 3. Surface representation (Nicholls *et al.*, 1993) of coagulogen and a putative coagulin monomer lacking the peptide C fragment at the top. The surface is colour coded with the electrostatic potential: blue, positively charged regions; red, negatively charged regions. A distinct hydrophobic cove becomes accessible upon releasing the peptide C fragment. It is flanked by Glu155 and Arg115, leading to a distinctive charge–hydrophobic–charge pattern. A similar pattern can be observed on other surface regions. 'N' indicates the N-terminus.

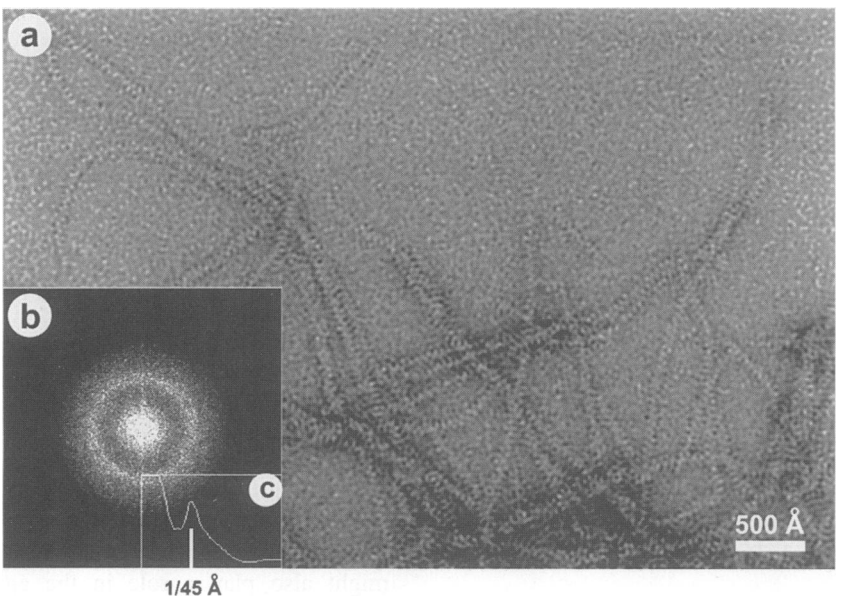

Fig. 4. (**a**) Coagulin network. When reacting with trypsin or clotting enzyme, coagulogen is transformed into a network consisting of linear fibres with branching points at irregular distances. In order to improve the visualization of the fibre structure, a low pass filter cutting off spatial frequencies above (1/39)Å^{-1} was applied. (**b**) Fourier spectrum of (a). The ring in the spectrum indicates a (poorly defined) periodicity of ~45 Å. (**c**) Azimuthally averaged Fourier spectrum showing broadening of the 45 Å periodicity.

some parts of the images, single fibres can be seen, suggesting that the fibres may have a helical structure with a diameter of ~100 Å, a pitch of 45 Å and an ~180° increment between repeating AB structural units. However, the helix seems to be flexible, and the broad ring of the power spectrum suggests a greater variability in the aggregation of the AB fragments. This could, at least partially, explain the thickness variability of the fibres. There also seems to be a tendency for single fibres to aggregate laterally, and thus form thicker fibres.

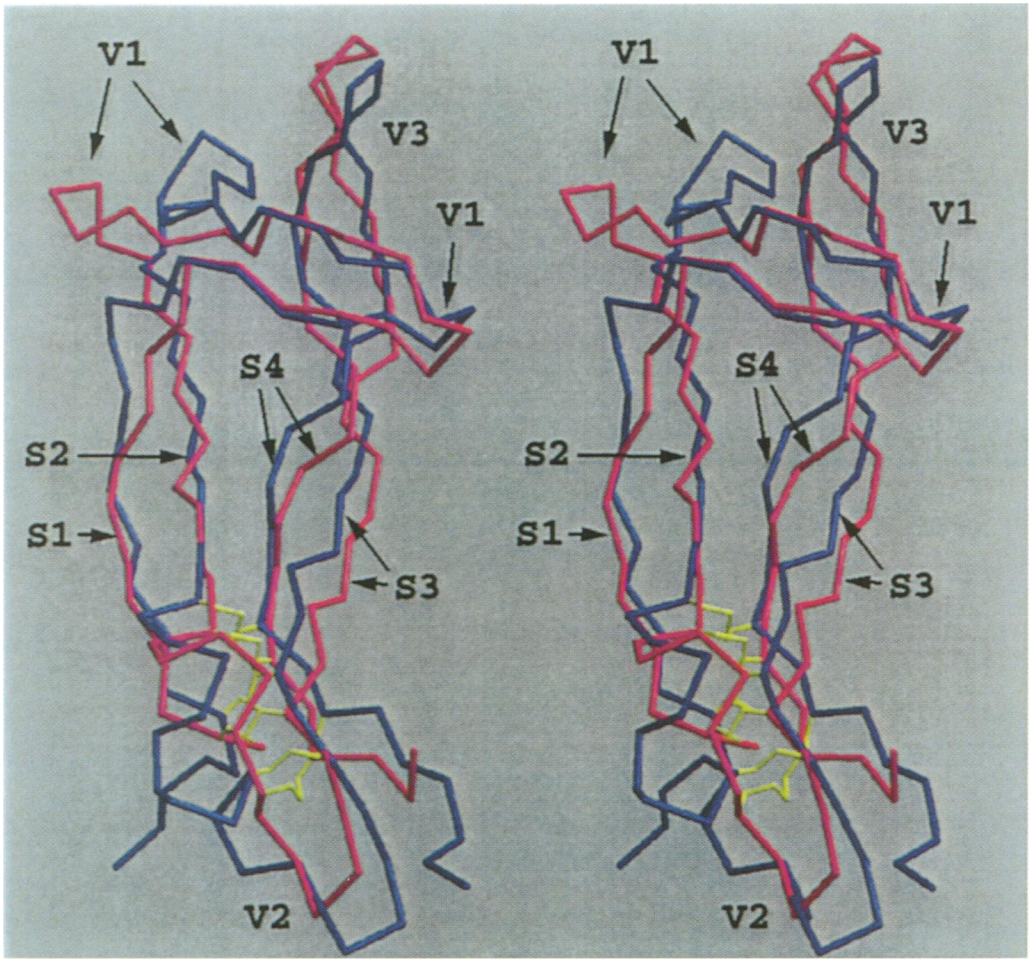

Fig. 5. Superposition of the C_α traces of coagulogen BII domain (red) and NGF (blue). S1–S4 (B1, B4, B5 and B6 in coagulogen) indicate β-strands common to all members of the cystine knot superfamily. In addition, the variable regions V1–V3 inserted in this basic motif are quite similar for coagulogen and NGF. V1 consists of the β-sheet β3 (B2 and B3 in coagulogen) across the molecule axis and two loops (T23 and T34 in coagulogen). In NGF, the part corresponding to T34 is built up by a small additional β-sheet. V1–V3 are quite different for other members of the superfamily such as TGF-β, PDGF and hCG. The disulfide bridges involved in the cystine knot are depicted in yellow. Cys88–Cys140 in coagulogen corresponds to Cys15–Cys80 in NGF and represents a left hand spiral. A right hand hook type is given for the pair Cys127–Cys170 (coagulogen) and Cys58–Cys108 (NGF). The third pair is a right hand spiral: Cys134–Cys172 (coagulogen) corresponds to Cys68–Cys110 (NGF). The three common disulfide bridges reveal the same type of disulfide bridge.

Discussion

Coagulogen C-terminal domain has an NGF-like fold

The coagulogen structure has been compared with all protein structures deposited in the Brookhaven Protein Database using the program GBF3DFIT (Lessel and Schomburg, 1994). A striking topologic similarity of the C-terminal half domain of coagulogen, BII, to NGF (McDonald *et al.*, 1991), a member of the neurotrophin family, was found (see Figure 5). The three β-sheets β1 (B5 and B6), β2 (B1 and B4) and β3 (B2 and B3) of coagulogen's BII domain correspond to β-sheets with identical topology in NGF. The loops between them are similar, and also the unusually large twist of the β-sheet is present in NGF. A sequence alignment based on topological equivalence (Figure 6) shows a sequence identity of 20.7% in the topologically equivalent regions. In NGF, the six cysteines exhibit a characteristic disulfide connectivity pattern (1–4, 2–5, 3–6), with the (1–4) disulfide bridge penetrating a macrocyclic ring formed by the disulfide bridges (2–5) and (3–6) and the polypeptide chain. This folding motif has been termed a cystine knot (McDonald and Hendrickson, 1993). Six of the cysteines in coagulogen (1–4: 88–140, 2–5: 127–170, 3–6: 134–172) are interconnected in an identical manner, assigning coagulogen to be a cystine knot protein. The r.m.s. deviation between all atoms of these six common cysteines in coagulogen and NGF is 0.83 Å (0.55 Å only for the C_α atoms).

Coagulogen belongs to the cystine knot superfamily

Several protein structures containing the cystine knot motif and sharing a common overall topology have been eludicated, making up a superfamily of structurally related proteins (McDonald and Hendrickson, 1993; Murray-Rust *et al.*, 1993; Sun and Davies, 1995; Isaacs, 1995). This cystine knot superfamily has been divided into several subclasses: the neurotrophins, the TGF-β (transforming growth factor-β) family, the PDGF (platelet-derived growth factor) family and the glycoprotein hormones. At least one member of each subclass has been characterized

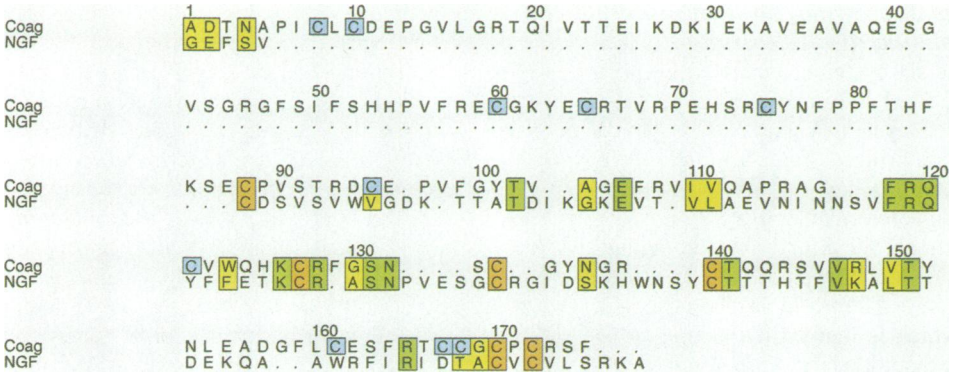

Fig. 6. Sequence alignment (Barton, 1993) of coagulogen and NGF (Lessel and Schomburg, 1994), based on the knowledge of the structure. The six cysteines forming the disulfide knot motif of the TGF-β2 superfamily are conserved (orange). The most significant similarity between the sequences could be observed for two triplets of identical residues (118–120 and 126–128, shown in green), and nearby the cysteines 140, 170 and 172. Similar residues are depicted in yellow. The cysteines specific for coagulogen are depicted in blue. There is no evidence for a functional meaning of the conserved regions, e.g. receptor binding of NGF.

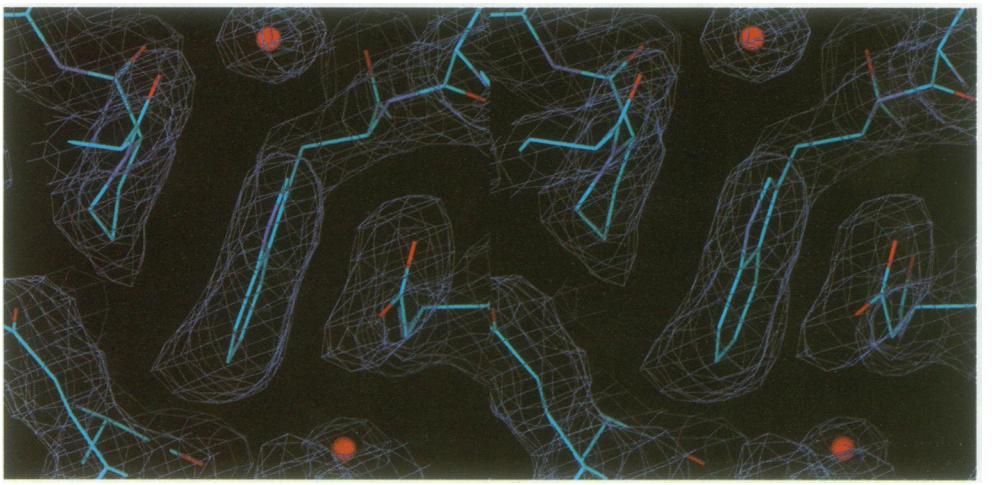

Fig. 7. Representative part of the final $2F_o–F_c$ density map, contoured at the 1σ level. The figure shows the only tryptophan of the structure, Trp123 in the BII fragment. Red balls represent water molecules.

structurally. NGF, which belongs to the neurotrophins, was the first solved structure exhibiting the cystine knot (McDonald *et al.*, 1991). In the TGF-β subclass, the structures of TGF-β2 (Daopin *et al.*, 1992, 1993) and, very recently, of TGF-β1 (Hinck *et al.*, 1996) and TGF-β3 (Mittl *et al.*, 1996) have been solved. Furthermore, the structures of one homodimeric form of PDGF, human PDGF-BB (Oefner *et al.*, 1992), and the glycoprotein hormone human chorionic gonadotrophin (hCG) (Laptorn *et al.*, 1994; Wu *et al.*, 1994) have been solved. Further proteins have been proposed to be members of the cystine knot superfamily (Meitinger *et al.*, 1993; McDonald and Chao, 1995; Noguti *et al.*, 1995). All members of the cystine knot superfamily currently known typically form homo- and heterodimers. Each representative of known structure has a distinctive subunit arrangement and inter-subunit bonding pattern within its dimer.

While the neurotrophins, the TGF-β proteins, the PDGFs and the glycoprotein hormones act on a variety of different cells, they all function by inducing a biological response allowing binding to specific receptors. Coagulogen, in contrast, differs in its mechanism of action from the currently known members of the cystine knot superfamily.

All members of the cystine knot motif superfamily consist of four conserved β-strands S1–S4 (in coagulogen the β-strands B1, B4, B5 and B6), which form two antiparallel β-sheets (S1 and S2, S3 and S4). Those β-strands are interconnected by the three disulfide bridges of the cystine knot. In this common motif, different variable regions V1–V3 are inserted, giving rise to the variety of the subclasses of the cystine knot superfamily. In contrast to TGF-β, PDGF and hCG, coagulogen and NGF show almost identical variable regions V1–V3. V1 is formed by the β-sheet β3 (strands B2 and B3) and its adjacent turns (T12, T23 and T34). V2 and V3 are equivalent to the hairpin turns T45 and T56 (Figure 5). Furthermore, NGF and coagulogen exhibit the same nine residue spacing between the knot cysteines 2 and 3. In contrast, TGF-β, PDGF and hCG show a three residue spacing between those cysteines (the spacing between the knot cysteines 5 and 6 is one residue for all discussed members of the cystine knot superfamily). Therefore, the structures of the coagulogen BII domain and NGF are more closely related compared with the other subclasses of the cystine knot superfamily. However, due to its quite different function, coagulogen should be considered to be a member of a new subclass of the cystine knot superfamily.

The N-terminus of coagulogen is anchored to the NGF-like domain via disulfide bridges

In addition to the six cysteines which form the conserved cystine knot motif of the superfamily, coagulogen possesses 10 further cysteines that are involved in five additional disulfide bridges. Two and three of these disulfides anchor fragment A and fragment BI to the NGF-like BII domain, respectively. Compared with the growth factors, the additional 87 amino acids contained in fragments A, C and BI render the coagulogen molecule more bulky and provide a distinct hydrophobic core region in its upper part, which is flanked by its unique N-terminal half domain. The disulfide bridges between the BII domain and fragments BI and A seem to be vital for the stabilization of the whole N-terminal half of coagulogen, as this part, except for the peptide C helix, reveals mainly multiple coil structures. The structural importance of these disulfide bridges for stabilizing the coagulogen molecule is also indicated by the strong conservation of all 16 cysteines in the coagulogens of all four living *Limulidae* species (Iwanaga *et al.*, 1986).

Dimerization of the cystine knot proteins

The active forms of all growth factors of the cystine knot superfamily are dimers. Each subclass of the cystine knot superfamily exhibits a distinctive mode of dimerization, given by the relative arrangement of the protomers and the type of binding (McDonald and Hendrickson, 1993; Sun and Davies, 1995). Non-covalent linkage of the monomers is seen in the case of NGF and hCG, with NGF forming homodimers, and hCG heterodimers. In TGF-β and PDGF proteins, the protomers are linked by one and two additional interprotomeric disulfide bridges.

In contrast to these proteins, dimerization of coagulogen does not seem to be necessary for its function. However, it may be considered as a new dimerization mode of the cystine knot proteins seen thus far only in crystal packing. The interface between both coagulogen monomers in the crystals is made from the BI segment in protomer 1 and the peptide C helix in protomer 2, and has a size of ~890 Å^2. There is an intraprotomeric salt bridge formed by the side chains of Arg69 and Glu35. In contrast to the growth factors, the NGF-like domain BII, which is almost covered by the N-terminal half, is not involved in dimer formation.

Parallel between horseshoe crab and Drosophila proteinase cascades

An interesting parallel between the *Limulus* clotting cascade and a proteinase cascade in *Drosophila* has been detected recently. The horseshoe crab clotting enzyme and the *Drosophila* serine proteinases Snake and Easter show significant sequence homology and probably identical disulfide connectivity patterns in their proparts (Gay and Keith, 1992; Smith and DeLotto, 1992). They have therefore been considered as members of a distinct subfamily of regulating serine proteinases. *Limulus* factor B also exhibits a similar primary structure and disulfide linkage to the clotting enzyme and Easter (Muta *et al.*, 1993a). Snake and Easter are members of a proteinase cascade that is involved in a signal transduction pathway determining the dorsal–ventral cell fate during *Drosophila* embryogenesis (Smith and DeLotto, 1994). In this cascade, activated Snake leads to activation of Easter, which in turn converts Spätzle (Stein and Nüsslein-Volhard, 1992; Smith and DeLotto, 1994) into the presumed ligand that activates the Toll receptor (Stein *et al.*, 1991; Roth, 1994). Interestingly, it has been noted that the special arrangement of cysteines in the C-terminal portion of Spätzle shows similarities with the cystine knot motif first found in NGF (McDonald *et al.*, 1991; Casanova *et al.*, 1995). Recently, DeLotto *et al.* have provided experimental evidence suggesting that the cystine connectivity pattern of the C-terminal part of Spätzle is like that observed in NGF and that the protein conformations might be similar (R.DeLotto, personal communication). Thus, the NGF-like BII domain of coagulogen likewise seems to have a structure closely related to Spätzle. The unexpected structural similarities between the target proteins of the cascades in horseshoe crab and *Drosophila*, coagulogen and Spätzle, and the proposed similarity between at least two of their processing serine proteinases in each cascade (factor B, proclotting enzyme versus Snake and Easter) strongly suggest a common ancestry of the two serine proteinase cascades in the two invertebrate families. The finding of a protein with a neurotrophin fold in a living fossil such as the horseshoe crab demonstrates that this folding motif is much older than the vertebrates, and is particularly interesting in view of the fact that so far no neurotrophin-like molecules have been identified in invertebrates.

Materials and methods

Protein crystallization

Coagulogen was purified from Japanese horseshoe crab (*T.tridentatus*) haemocytes, collected from their haemolymph, as described previously (Iwanaga *et al.*, 1986). Crystals were grown by vapour diffusion using the hanging drop method at 20°C. Droplets mixed from 3.5 μl of coagulogen (10–12 mg/ml) and 2.5 μl of buffer (0.5 M ammonium sulfate, 0.1 M sodium acetate, 0.1 mM cadmium sulfate at pH 4.5) were concentrated against the reservoir buffer composed of the same components. The best crystals were obtained by macroseeding, by adding tiny, well grown crystals to freshly prepared droplets. The rhombohedral shaped crystals typically grew to a size of 0.8×0.5×0.4 mm^3 within 7 days. They belong to the space group $P3_221$ having unit cell parameters of $a = b = 52.2$ Å, $c = 232.2$ Å, $\alpha = \beta = 90°$, $\gamma = 120°$. The crystals contain two molecules per asymmetric unit. The water content of the crystals is 47% ($V_m = 2.33$).

X-ray data collection, structure determination and refinement

Data were collected using Cu_α radiation from a Rigaku rotating anode X-ray generator operated at 4.8 kW as well as using synchrotron radiation at the beamline BW6 at DESY (Hamburg). All data were recorded on 22 cm MAR Research image plate detectors, and were processed with MOSFLM (Leslie, 1991), ROTAVATA, AGROVATA and TRUNCATE (Collaborative Computing Project Number 4, 1994). Native data were collected from one crystal. Data collection statistics are summarized in Table I. An extensive search for heavy atom derivatives led to four suitable derivatives: crystals soaked with $K_2[ReCl_6]$ (RECL), $K_2[PtI_6]$ (PTI6), platinum(II) 2,2′-6′,2″-ter-pyridinium chloride (PTPY) and sodium-bis-(*N*-methyl-hydantoino)gold(I) (AUHY) provided clearly interpretable isomorphous heavy atom derivatives. Simultaneous soaking with $K_2[ReCl_6]$ and $K_2[PtI_6]$ led to a reasonable double derivative (REPT). In addition to derivative data collection in-house, data from two derivatives, $K_2[ReCl_6]$ (RESY) and $K_2[PtI_6]$ (PTSY), were also collected at the synchrotron (DESY) using wavelengths near the appropriate absorption edges in order to get larger anomalous effects for phasing. The derivative data were interpreted by calculating difference Patterson, anomalous Patterson and vector verification maps. The heavy atom positions obtained were judged by calculating cross-phased difference Fourier maps using the program package PROTEIN (Steigemann, 1974). Refinement of heavy atom positions and phasing was performed

Table I. Data collection and refinement statistics

Parameter	NATI	RECL	RESY	PTPY	PTI6	PTSY	AUHY	REPT
Data collection								
Wavelength (Å)	1.0	Cu_α	0.96	Cu_α	Cu_α	0.99	Cu_α	Cu_α
Resolution (Å)	2.0	2.9	2.1	2.8	3.2	2.3	2.7	2.8
Completeness (%)[a]	94.8/89.3	98.3/90.6	96.7/83.5	98.0/83.4	99.3/96.0	93.7/87.8	93.3/95.1	96.6/93.0
Measurements	44 956	26 680	79 692	20 227	12 592	58 270	33 210	20 799
Uniques	24 419	8602	22 063	9620	6110	16 648	9852	9204
MIRAS phasing								
R_{merge}[b](%)	4.2	3.5	3.3	5.2	4.0	4.0	5.0	5.7
R_{iso}[c](%)		18.3	16.5	14.8	21.0	16.1	19.9	22.3
R_{cullis}[d]		0.61	0.62	0.95	0.85	0.88	0.92	0.70
Phasing power[e]		2.05	1.96	0.49	0.93	0.79	0.62	1.59
Sites		1	1	1	1	1	2	2
FOM	0.72							

[a]Overall/last resolution shell.
[b]$R_{merge} = \Sigma\Sigma|I(h)_i - \langle I(h)\rangle|/\Sigma\Sigma I(h)_i$.
[c]$R_{iso} = \Sigma|F_{PH} - F_P|/\Sigma F_P$.
[d]$R_{cullis} = \Sigma||F_{PH} \pm F_P|/\Sigma|F_{PH} - F_P|$.
[e]$\langle F_H\rangle/\langle E\rangle$.

with MLPHARE (Otwinowski, 1991; Collaborative Computing Project Number 4, 1994). Phasing statistics are summarized in Table I. Unambiguous determination of the correct space group $P3_221$ and handedness were performed by refining the anomalous occupancies in space groups $P3_121$ and $P3_221$, using two derivatives with large anomalous effects (RESY and PTSY). The initial model building was performed by fitting polyalanine fragments manually to the MIRAS-phased 3.0 Å electron density map using the programs MAIN (Turk, 1992) and O (Jones *et al.*, 1991). The model was completed by successive cycles of refinement, consisting of phase combination of MIRAS and partial model phases with SIGMAA (Read, 1986), simulated annealing and Powell minimization methods using XPLOR (Brünger, 1992), and remodelling at the display. In the refinement, the force field parameters from Engh and Huber (1991) were used.

From an estimation of the packing density (Matthews, 1968) (V_m = 2.3 for a dimer, 4.6 for a monomer), we originally assumed a dimer in the asymmetric unit. The finding of single sites in all rhenium and platinum derivatives, the absence of pseudo origin peaks in the native Patterson map and the absence of any significant solutions in the self-rotational searches performed with XPLOR (Brünger, 1992), GLRF (Tong and Rossmann, 1990) and AMoRe (Navazza, 1987) led us to suppose a monomer in the asymmetric unit [several protein crystals with unusually high V_m values have been reported (Bode and Schirmer, 1985)]. However, by measuring the density of the crystals as described by Kiefersauer *et al.* (1996), the existence of a dimer was clearly established. Subsequent structural analysis showed protection of binding sites for Pt (His54) and Re (Glu87) in one of the protomers. A reasonable part of its surface is covered by the second protomer. Two-fold NCS averaging of the electron density was performed using the RAVE-MAMA package (Kleywegt and Jones, 1993) only after enough identical segments could be identified and assigned to the two distinct molecules. The final refinement was carried out with no NCS constraints. Comparing the two protomeric molecules, the course of the backbone differs in the region around Ile50 (BI), which is involved in dimer formation, giving rise to a change of the backbone course for residues Phe48–His54. The r.m.s. deviation of C_α atoms between the monomers is 0.34 Å considering all C_α atoms except those of regions involved in dimer formation or loop regions. In the final $2F_o$–F_c map (as well as in the initial MIRAS map), only poor electron density at best could be observed (in both molecules) for the following regions: Ala1–Pro6, Pro13–Thr23, Gly42–Gly47. The current model was refined with all data between 2.0 and 6.0 Å resolution (24 371 unique reflections) to a crystallographic *R*-factor of 19.1% and a free *R*-value R_f of 28.2%. It comprises 2740 non-hydrogen protein atoms, 195 solvent molecules interpreted as water and one sulfate molecule. The r.m.s. deviations from target values of bonds and angles are 0.012 Å and 2.2°, respectively. The r.m.s. deviation of the *B*-factors of bonded atoms is 2.55 Å^2.

Electron microscopy

For electron microscopy preparation, a 9 µl solution of coagulogen (1 mg/ml in 20 mM Tris–HCl, pH 7.0) and a 1 µl solution of either trypsin or clotting enzyme from Japanese horseshoe crab (0.1 mg/ml in 1 mM HCl) were applied to a copper grid that was coated with carbon film. After a reaction time in the range of 15–120 s, the liquid was blotted with filter paper, and the specimen negatively stained with a 2% solution of uranyl acetate. Specimens were investigated in a CM 10 transmission electron microscope (Philips) at 100 kV accelerating voltage. Images were recorded at magnifications of 6600× and 39 000×. As a control, the same preparation was carried out without addition of trypsin or clotting enzyme. In this case, no network structure could be found.

Structural analysis

A detailed secondary structure assignment was performed using the program STRIDE (Frishman and Argos, 1995). STRIDE found essentially the same but more extended β-strands compared with DSSP (Kabsch and Sander, 1983). A β-bulge caused by an additional residue (Cys168) in the β1 sheet divides the β1 sheet into two regular β-sheets. However, we consider the total segment as β-sheet. The results of STRIDE were used for secondary structure assignment and the ribbon plot (Figures 1), neglecting the β-bulges in the β1 sheet. Further structural analysis, in particular analysis of β-bulges and disulfide bridges, was done using the program PROMOTIF (Hutchinson and Thornton, 1996).

Acknowledgements

We thank I.Mayr for preliminary crystallization experiments; Dr Takeshi Shigenaga for cooperating on the purification of coagulogen from the haemocyte lysate; Professor R.Jaenicke for confirmation of the monomer status of coagulogen at high protein concentration by ultracentrifugation; Professor Y.A.Barde for valuable discussions and helpful suggestions; Drs H.Bartunik, M.Kobarg and G.Buth for help with X-ray data collection at the synchrothron of DESY; Drs T.Mather and M.T.Stubbs for careful and critical reviewing of this manuscript; and Dr R.DeLotto for providing results about the proposed topological similarity of the C-terminus of Spätzle and NGF before publication. Part of this work was supported by a Grant in Aid from the Japan Society for the Promotion of Science (JSPS).

References

Barton,G.J. (1993) ALSCRIPT: a tool to format multiple sequence alignments. *Protein Engng*, **6**, 37–40.

Bode,W. and Schirmer,T. (1985) Determination of the protein content of crystals formed by *Mastigocladus laminosus* C-phycocyanin, *Chroomonas* spec., phycocyanin-645 and modified human fibrinogen using an improved Ficoll density gradient method. *Biol. Chem. Hoppe-Seyler*, **366**, 287–295.

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

Casanova,J., Furriols,M., McCormick,C.A. and Struhl,G. (1995) Similarities between trunk and spatzle, putative extracellular ligands specifying body pattern in *Drosophila*. *Genes Dev.*, **9**, 2539–2544.

Collaborative Computing Project Number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr.*, **D50**, 760–763.

Daopin,S., Piez,K.A., Ogawa,Y. and Davies,D.R. (1992) Crystal structure of transforming growth factor-β2: an unusual fold for the superfamily. *Science*, **257**, 369–373.

Daopin,S., Li,M. and Davies,D.R. (1993) Crystal structure of TGF-β2 refined at 1.8 Å resolution. *Protein Struct. Funct. Genet.*, **17**, 176–92.

Engh,R.A. and Huber,R. (1991) Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallogr.*, **A47**, 392–400.

Frishman,D. and Argos,P. (1995) Knowledge-based protein secondary structure assignment. *Proteins*, **23**, 566–579.

Gay,N.J. and Keith,F.J. (1992) Regulation of translation and proteolysis during the development of embryonic dorso-ventral polarity in *Drosophila*. Homology of easter proteinase with *Limulus* proclotting enzyme and translational activation of Toll receptor synthesis. *Biochim. Biophys. Acta*, **1132**, 290–296.

Hinck,A.P. *et al.* (1996) Transforming growth factor β1: three-dimensional structure in solution and comparison with the X-ray structure of transforming growth factor β2. *Biochemistry*, **35**, 8517–8534.

Hutchinson,E.G. and Thornton,J.M. (1996) PROMOTIF—a program to identify and analyse structural motifs in proteins. *Protein Sci.*, **5**, 212–220.

Isaacs,N.W. (1995) Cystine knots. *Curr. Opin. Struct. Biol.*, **5**, 391–395.

Iwanaga,S. (1993a) Primitive coagulation systems and their message to modern biology. *Thromb. Haemostasis*, **70**, 48–55.

Iwanaga,S. (1993b) The *Limulus* clotting reaction. *Curr. Opin. Immunol.*, **5**, 74–82.

Iwanaga,S., Morita,T., Miyata,T., Nakamura,T. and Aketagawa,J. (1986) The hemolymph coagulation system in invertebrate animals. *J. Protein Chem.*, **5**, 255–268.

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

Kabsch,W. and Sander,C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.

Kiefersauer,R., Stetefeld,J., Gomis-Rüth,F.X., Romao,M.J., Lottspeich,F. and Huber,R. (1996) Protein-crystal density by volume measurement and amino-acid analysis. *J. Appl. Crystallogr.*, **29**, 1–7.

Kleywegt,G.J. and Jones,T.A. (1993) Masks made easy. *ESF/CCP4 Newslett.*, **28**, 56–59.

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

Laptorn,A.J., Harris,D.C., Littlejohn,A., Lustbader,J.W., Canfield,R.E., Machin,K.J., Morgan,F.J. and Isaacs,N.W. (1994) Crystal structure of human chorionic gonadotrophin. *Nature*, **369**, 455–461.

Leslie,A.G.W. (1991) Macromolecular data processing. In Moras,D., Podjarni,A.D. and Thierry,J.C. (eds), *Crystallographic Computing*. Oxford University Press, Oxford, pp. 50–61.

Lessel,U. and Schomburg,D. (1994) Similarities between protein 3-D structures. *Protein Eng.*, **7**, 1175–1187.

Levin,J. and Bang,F.B. (1964) The role of endotoxin in the extracellular coagulation of *Limulus* blood. *Bull. Johns Hopkins Hosp.*, **115**, 265–274.

Matthews,B.W. (1968) Solvent content of protein crystals. *J. Mol. Biol.*, **33**, 491–497.

McDonald,N.Q. and Chao,M.V. (1995) Structural determinants of neurotrophin action. *J. Biol. Chem.*, **270**, 19669–19672.

McDonald,N.Q. and Hendrickson,W.A. (1993) A structural superfamily of growth factors containing a cystine knot motif. *Cell*, **73**, 421–424.

McDonald,N.Q., Lapatto,R., Murray-Rust,J., Gunning,J., Wlodawer,A. and Blundell,T.L. (1991) New protein fold revealed by a 2.3 Å resolution structure of nerve growth factor. *Nature*, **354**, 411–414.

Meitinger,T., Meindl,A., Bork,P., Rost,B., Sander,C., Haasemann,M. and Murken,J. (1993) Molecular modelling of the Norrie disease protein predicts a cystine knot growth factor tertiary structure. *Nature Genet.*, **5**, 376–380.

Mittl,P.R.E., Priestle,J.P, Cox,D.A., McMaster,G., Cerletti,N. and Grütter,M.G. (1996) The crystal structure of tgf-β3 and comparison to tgf-β2: implications for receptor binding. *Protein Sci.*, **5**, 1261–1271.

Murray-Rust,J., McDonald,N., Blundell,T.L., Hosang,M., Oefner,C., Winkler,F. and Bradshaw,R.A. (1993) Topological similarities in TGF-β2, PDGF-BB and NGF define a superfamily of polypeptide growth factors. *Structure*, **1**, 153–159.

Muta,T. and Iwanaga,S. (1996) Clotting and immune defense in limulidae. In Rinkevich,B. and Müller,W.E.G. (eds), *Progress in Molecular and Subcellular Biology: Invertebrate Immunology*. Springer Verlag, Heidelberg, Vol. 15, pp. 154–189.

Muta,T., Oda,T. and Iwanaga,S. (1993a) Horseshoe crab coagulation factor B. *J. Biol. Chem.*, **268**, 21383–21388.

Muta,T., Tokunaga,F., Nakamura,T., Morita,T. and Iwanaga,S. (1993b) *Limulus* clotting factor C: lipopolysaccharide-sensitive serine protease zymogen. *Methods Enzymol.*, **223**, 336–345.

Muta,T., Seki,N., Takaki,Y., Hashimoto,R., Oda,T., Iwanaga,A., Tokunaga,F. and Iwanaga,S. (1995) Purified horseshoe crab factor G. Reconstitution and characterization of the (1→3)-β-D-glucan-sensitive serine protease cascade. *J. Biol. Chem.*, **270**, 892–897.

Nakamura,S., Iwanaga,S., Harada,T. and Niwa,M. (1976) A clottable protein (coagulogen) from amoebocyte lysate of Japanese horseshoe crab (*Tachypleus tridentatus*): its isolation and biochemical properties. *J. Biochem.*, **80**, 1011–1021.

Navazza,J. (1987) On the fast rotation function. *Acta Crystallogr.*, **A43**, 645–653.

Nicholls,A., Bharadwaj,R. and Honig,B. (1993) Grasp: graphical representaion and analysis of surface properties. *Biophys. J.*, **64**, A166.

Noguti,T. *et al.* (1995) Insect prothoracicotropic hormone: a new member of the vertebrate growth factor superfamily. *FEBS Lett.*, **376**, 251–256.

Oefner,C., D'Arcy,A., Winkler,F.K., Eggimann,B. and Hosang,M. (1992) Crystal structure of human platelet-derived growth factor BB. *EMBO J.*, **11**, 3921–3926.

Otwinowski,Z. (1991) MLPHARE. CCP4 Proceedings 80–88. Daresbury Laboratory, Warrington, UK.

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

Roth,S. (1994) Proteolytic generation of a morphogen. *Curr. Biol.*, **4**, 755–757.

Smith,C.L. and DeLotto,R. (1992) A common domain within the proenzyme regions of the *Drosophila* snake and easter proteins and *Tachypleus* proclotting enzyme defines a new subfamily of serine proteases. *Protein Sci.*, **1**, 1225–1226.

Smith,C.L. and DeLotto,R. (1994) Ventralizing signal determined by protease activation in *Drosophila* embryogenesis. *Nature*, **368**, 548–551.

Srimal,S., Miyata,T., Kawabata,S., Miyata,T. and Iwanaga,S. (1985) The complete amino acid sequence of coagulogen isolated from southeast Asian horseshoe crab, *Carcinoscorpius rotundicauda*. *J. Biochem.*, **98**, 305–318.

Steigemann,W. (1974) Die entwicklung und anwendung von rechenverfahren und rechenprogrammen zur strukturanalyse von proteinen am beispiel des trypsin-trypsin inhibitor komplexes, des freien inhibitors und der L-asparagine. Ph.D. thesis, TU München, Germany.

Stein,D. and Nüsslein-Volhard,C. (1992) Multiple extracellular activities in *Drosophila* egg perivitelline fluid are required for establishment of embryonic dorsal–ventral polarity. *Cell*, **68**, 429–440.

Stein,D., Roth,S., Vogelsang,E. and Nüsslein-Volhard,C. (1991) The polarity of the dorsoventral axis in the *Drosophila* embryo is defined by an extracellular signal. *Cell*, **65**, 725–735.

Sun,P.D. and Davies,D.R. (1995) The cystine-knot growth-factor superfamily. *Annu. Rev. Biophys. Biomol. Struct.*, **24**, 269–291.

Tanaka,S. and Iwanaga,S. (1993) *Limulus* test for detecting bacterial endotoxins. *Methods Enzymol.*, **223**, 358–364.

Tong,L. and Rossmann,M.G. (1990) The locked rotation function. *Acta Crystallogr.*, **A46**, 783–792.

Turk,D. (1992) Weiterentwicklung eines programmes für molekülgraphik und elektronendichte-manipulation und seine anwendungen auf verschiedene protein-strukturaufklärungen. Ph.D. thesis, TU München, Germany.

Wu,H., Lustbader,J.W., Liu,Y., Canfield,R.E. and Hendrickson,W.A. (1994) Structure of human chorionic gonadotropin at 2.6 Å resolution from MAD analysis of the selenomethionine protein. *Structure*, **2**, 545–558.

Received on August 5, 1996; revised on September 12, 1996