# Crystal structure of a 30 kDa C-terminal fragment from the γ chain of human fibrinogen

Vivien C Yee[1,2], Kathleen P Pratt[1], Hélène CF Côté[1], Isolde Le Trong[2,3], Dominic W Chung[1], Earl W Davie[1], Ronald E Stenkamp[2,3] and David C Teller[1,2]*

**Background:** Blood coagulation occurs by a cascade of zymogen activation resulting from minor proteolysis. The final stage of coagulation involves thrombin generation and limited proteolysis of fibrinogen to give spontaneously polymerizing fibrin. The resulting fibrin network is covalently crosslinked by factor XIIIa to yield a stable blood clot. Fibrinogen is a 340 kDa glycoprotein composed of six polypeptide chains, $(\alpha\beta\gamma)_2$, held together by 29 disulfide bonds. The globular C terminus of the γ chain contains a fibrin-polymerization surface, the principal factor XIIIa crosslinking site, the platelet receptor recognition site, and a calcium-binding site. Structural information on this domain should thus prove helpful in understanding clot formation.

**Results:** The X-ray crystallographic structure of the 30 kDa globular C terminus of the γ chain of human fibrinogen has been determined in one crystal form using multiple isomorphous replacement methods. The refined coordinates were used to solve the structure in two more crystal forms by molecular replacement; the crystal structures have been refined against diffraction data to either 2.5 Å or 2.1 Å resolution. Three domains were identified in the structure, including a C-terminal fibrin-polymerization domain (P), which contains a single calcium-binding site and a deep binding pocket that provides the polymerization surface. The overall structure has a pronounced dipole moment, and the C-terminal residues appear highly flexible.

**Conclusions:** The polymerization domain in the γ chain is the most variable among a family of fibrinogen-related proteins and contains many acidic residues. These residues contribute to the molecular dipole moment in the structure, which may allow electrostatic steering to guide the alignment of fibrin monomers during the polymerization process. The flexibility of the C-terminal residues, which contain one of the factor XIIIa crosslinking sites and the platelet receptor recognition site, may be important in the function of this domain.

Addresses: [1]Department of Biochemistry, [2]Biomolecular Structure Center and [3]Department of Biological Structure, University of Washington, Seattle, WA 98195, USA.

*Corresponding author.
E-mail: teller@u.washington.edu

Key words: blood coagulation, calcium binding, factor XIIIa crosslinking, fibrinogen, fibrin polymerization

Received: 28 October 1996
Revisions requested: 25 November 1996
Revisions received: 4 December 1996
Accepted: 4 December 1996

Electronic identifier: 0969-2126-005-00125

Structure 15 January 1997, 5:125–138

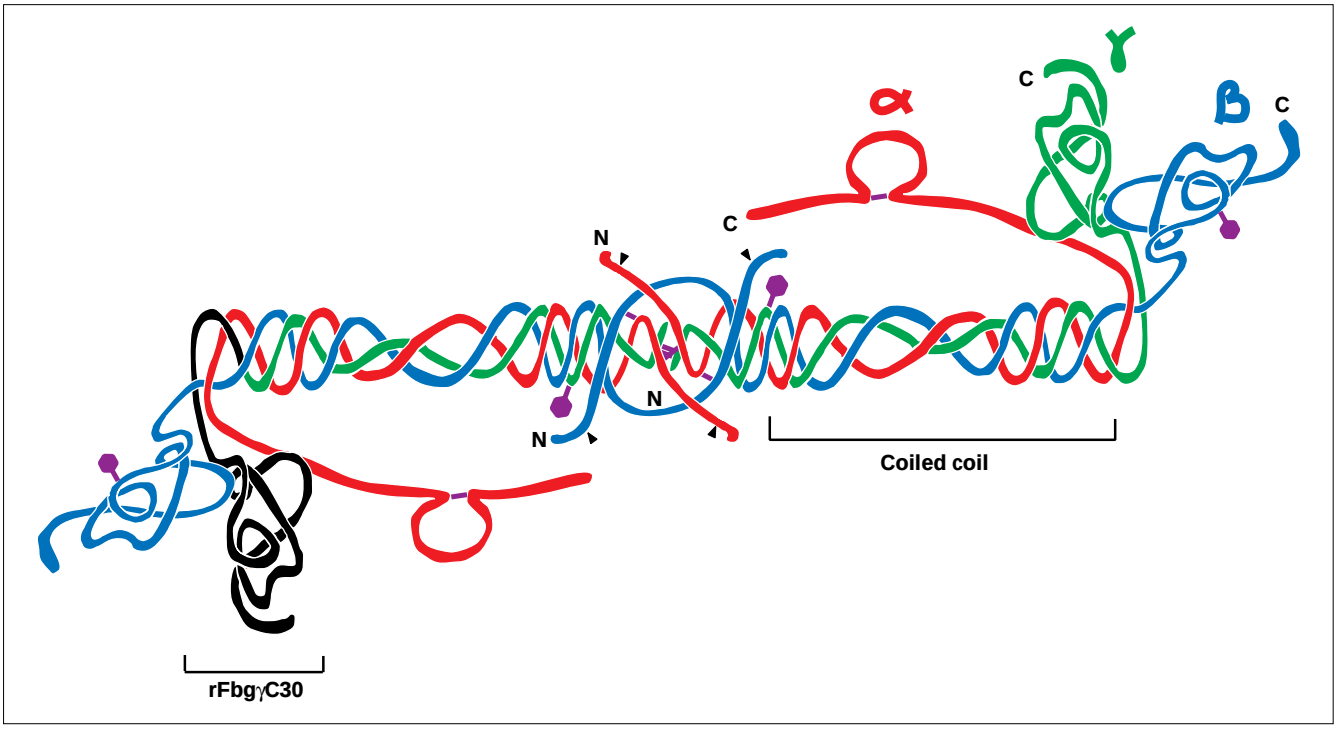


## Introduction

Fibrinogen is a large glycoprotein (molecular weight 340 KDa) circulating in blood. It is composed of six polypeptide chains $(\alpha\beta\gamma)_2$ arranged as two linked units, each containing an α, β and γ chain (Fig. 1). The protein appears as a trinodular structure in electron micrographs with a central nodule that contains the N-terminal regions of all six polypeptide chains [1,2]. The two distal nodules contain the C-terminal regions of the polypeptide chains and are joined to the central nodule by two separate triple-stranded coiled-coil regions [3,4]. The overall length of the molecule is 450 Å.

In the final stages of the coagulation cascade, fibrinogen is converted to fibrin by thrombin. In this reaction, fibrinopeptides are released from the N-termini of the two α chains, leading to rapid polymerization of the fibrin monomers. Fibrinopeptides are also released from the N-termini of the two β chains, but at a much slower rate. Protofibrils are generated initially as the fibrin monomers align in half-staggered and overlapping strands, showing a thickness of two molecules [5,6]. This alignment involves the newly generated N terminus of the α chain of one molecule binding to the C-terminal region of a γ chain of an adjacent molecule [7]. As the fibrin strands grow longer and thicker, the molecules are linked covalently by factor XIIIa, primarily through crosslinking of the γ chains [8,9]. The C-terminal region of the γ chain also participates in platelet plug formation, in which fibrinogen binds to and forms a bridge between activated platelets [10], during the aggregation phase of platelet plug formation. Fibrinogen contains three bound calcium ions [11], two of which are

**Figure 1**




Schematic representation of the fibrinogen molecule. The α, β, and γ chains are indicated, as well as their N and C termini. The purple hexagons represent N-linked glycosylations, while the small black triangles indicate the sites of thrombin cleavage. The region of the molecule encompassing the rFbgγC30 fragment is depicted in black. (Adapted from RF Doolittle.)

located in the C-terminal region of the γ chain; the third calcium ion is present in the central nodule [12]. Calcium accelerates fibrin formation [13], and also stabilizes the protein against proteolysis [14] and heat denaturation [15].

The goal of the present study was to establish the three-dimensional structure of the C-terminal region of the γ chain of fibrinogen by X-ray crystallography. A recombinant 30 kDa fragment from the C-terminal region of the human γ chain, which had been expressed in yeast, was isolated and crystallized for the structure determination. This fragment contains part of the polymerization surface of fibrin, a calcium-binding site, and the factor XIII cross-linking site. The results presented here demonstrate that this portion of the fibrin molecule possesses a unique fold, and the structural properties of the fragment provide several exciting insights into the mechanism of fibrin polymerization, the crosslinking of fibrin by factor XIIIa, platelet aggregation reactions, and calcium binding.

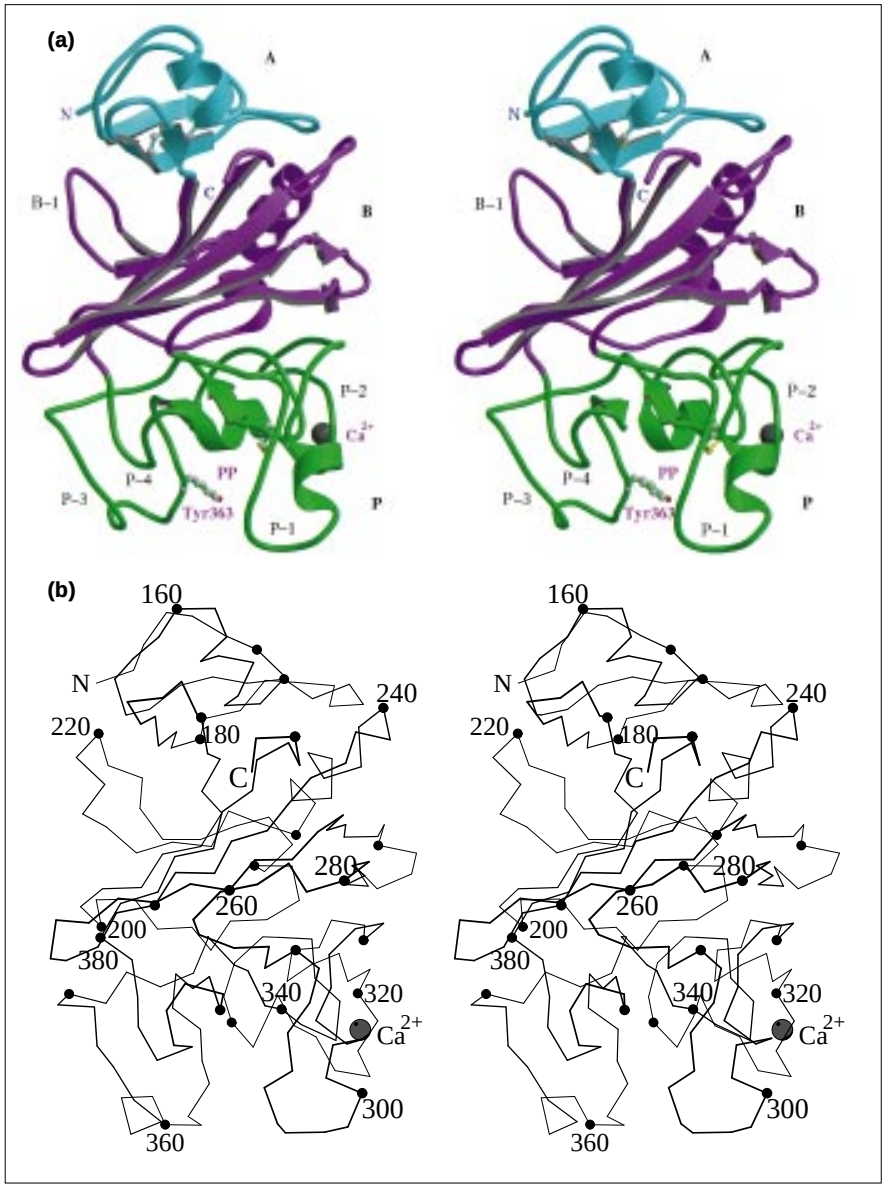
## Results and discussion

The three-dimensional structure of a recombinant 30 kDa C-terminal fragment of the human fibrinogen γ chain (rFbgγC30), containing residues Val143 to Val411 has been determined in three different crystal forms. The $P2_1$ crystal form was first solved by multiple isomorphous replacement methods. The resulting model, consisting of residues Gln144 to Leu392, one calcium ion, and 177 solvent molecules, has been refined against 10.0–2.1 Å resolution diffraction data to give a crystallographic R value of 15.9%. Two additional crystal forms with space groups R3 and $P2_12_12$ were solved by molecular replacement using the protein coordinates of the refined $P2_1$ structure. The R3 structure contains two independent molecules, the first model spans residues Gln144 to Gly403, and in the second residues Val143 to Leu392 and Gly397 to Gly403 are observed. The final R3 model, which includes one calcium ion per protein molecule and 202 solvent atoms, was refined against 10.0–2.5 Å resolution data to yield an R value of 17.1%. In the $P2_12_12$ crystal form, residues Gln144 to Leu402, one calcium ion, and 151 solvent molecules were observed and refined against 10.0–2.1 Å resolution diffraction data to give a final R value of 17.0%.

### Overall molecular structure

The rFbgγC30 fragment has a unique overall fold; no similar structure could be identified by using the computer program DEJAVU [16] or in a search of the Structural Classification of Proteins (SCOP) database [17]. The fragment is prolate ellipsoidal with dimensions of approximately 38×44×56 Å. There are three well defined

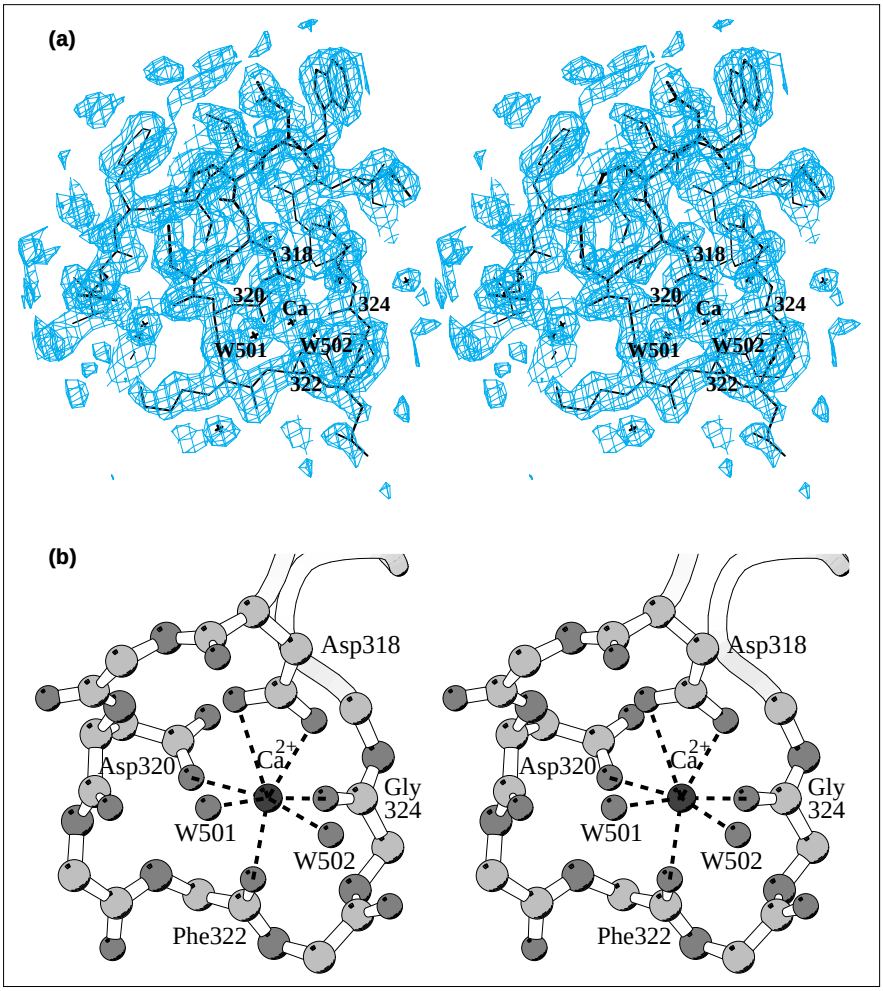
**Figure 2**




Structure of rFbgγC30 in the $P2_1$ crystal form. (a) The stereo ribbon diagram is colored by domains: A in cyan, B in purple, and P in green. The calcium ion is shown as a gray sphere; sidechains in the two disulfide bonds are drawn as ball-and-stick structures. Also labeled are the N and C termini of the chain, the polymerization pocket (PP) in the P domain, the sidechain of Tyr363 in the polymerization pocket, and the five variable loops (B-1, P-1 to P-4) described in the text. (b) Stereo diagram of the Cα trace with every 10th residue marked by a black sphere and every 20th residue labeled. The calcium ion is shown as a large gray sphere and labeled for reference. MOLSCRIPT [56] was used to generate (a) and (b) and Raster3D [57] was used to generate (a) only.

domains which pack sequentially in the globular body of the rFbgγC30 structure (Fig. 2). At the N terminus, the small A domain (Val143–Trp191) contains a twisted three-stranded antiparallel β sheet flanked by a short helix. As shown in Figure 2 this β structure consists of two sequential hairpin turns; the region of overlap, which forms the sheet, occurs only at residues Cys182 and Glu183. A disulfide bond (Cys153–Cys182) anchors the helix to the middle strand of the sheet. The second and central B domain (Thr192–Ala286 and Lys380–Leu392) contains a seven-stranded antiparallel β sheet with two short helices and a hairpin loop aligned against one of its faces. The two β strands of this sheet adjacent to the B-1 loop of Figure 2a (residues Phe215–His217 and Glu225–Trp227) are distorted relative to the other five strands of the sheet. A notable feature of the β sheet is that its middle strand is formed by the insertion of residues near the C terminus of the molecule (Lys381–Pro388). Little regular secondary structure is found in the third, or P, domain (Gly287–Met379), which is composed predominantly of loops. One of the two lobes in the P domain contains two short helices and the second disulfide bond (Cys326–Cys339) in the molecule. The preponderance of loops and the remarkably limited secondary structure in domain P are consistent with the high glycine content (13 of 94 amino acids) and suggest that this domain may be more flexible than domains A and B; atomic temperature factors indicate, however, that the A domain is the most mobile of the

**Figure 3**




The calcium-binding site in the $P2_1$ crystal form of rFbgγC30. (a) Stereo view of the $2F_o–F_c$ electron-density map calculated using phases from the final refined coordinates, contoured at 1σ level. The protein model is shown as thick bonds; the calcium ion, its protein ligands (residues 318, 320, 322, and 324) and two coordinating solvent molecules are labeled. (b) Stereo view of the calcium-binding loop P-2 with the protein mainchain, coordinating acidic sidechain groups, calcium ion, and bound solvent molecules drawn as ball-and-stick structures. Part (a) was produced by O [43] and part (b) was drawn using MOLSCRIPT [56].

three. In the $P2_1$ crystal form, the average temperature factor for the A, B, and P domains are 16.3, 9.6, and 12.1 Å$^2$, respectively. This pattern of average temperature factors for the three domains is similar in the R3 and $P2_12_12$ structures. The two reported plasmin cleavage sites (Lys302 and Lys356) [18] in the fragment are located in the P domain and have sidechains that are exposed to solvent and thus accessible to proteolytic cleavage.

### Calcium-binding site

Fibrin polymerization and crosslinking are calcium-dependent processes. A single calcium-binding site in the rFbgγC30 molecule has been identified in all three crystal forms, and coincides with the heavy atom position of the $TbCl_3$ derivative. The amino acid residues that bind the calcium ion are in a small loop (P-2) in the P domain (Fig. 3a,b), one end of which is connected to the rest of the domain by the Cys326–Cys339 disulfide bond. This disulfide bond has an important structural role, as revealed by the fact that Cys339 and Asn337 (which forms a mainchain hydrogen bond with Cys339) are the only two non-glycine residues in the structure that have glycine- like mainchain angles in the Ramachandran plot. Seven oxygen atoms coordinate the calcium ion in a geometry that is approximately octahedral. The oxygen atoms are provided by the sidechain carboxylate moieties of Asp318 and Asp320, along with the mainchain carbonyl groups of Phe322 and Gly324 and two solvent molecules (W501, W502) (Fig. 3b). Calcium to oxygen distances were not restrained during refinement. In the $P2_1$ structure, six of these distances are between 2.37 and 2.53 Å ,whereas the seventh (between the calcium and one of the two Asp318 carboxylate oxygen atoms) is longer, at 2.82 Å. This binding site is also the location of a defect in a congenitally abnormal fibrinogen, in which residues Asn319 and Asp320 are missing; this abnormal protein exhibits impaired calcium binding and fibrin polymerization [19]. Asp318, Asp320, and Gly324 are conserved among all the fibrinogen γ chain sequences. The conservation of Gly324 reflects the importance of the conformation of this loop for cation binding.

Equilibrium-dialysis binding studies have shown that the calcium-binding site encompasses the region of residues 303–355/356 [20]. Terbium fluorescence studies localized the calcium-binding site to residues 311–336 [21], consistent with our assignment of the protein ligands. A calcium-binding site has been predicted to be in this acidic region, but the predictions were for an EF-hand motif [21] or for a site similar to the cation site in the mannose-binding protein [18]. A comparison of the fibrinogen calcium-binding site with known EF-hand structures [22] and with rat mannose-binding protein [23] shows no notable similarity in the cation-binding sites, however. The short Asp318–Gly324 calcium-binding loop in rFbgγC30 differs significantly from all of the calcium binding structures currently accessible in the Protein Data Bank [22]. Thus, the rFbgγC30 structure adds one more to the wide variety of already known calcium-binding structural motifs [24].

#### Polymerization pocket

The conversion of fibrinogen to give spontaneously polymerizing fibrin is initiated when thrombin cleaves the fibrinopeptide A from each α chain to expose new N termini. These new termini then bind to the C-terminal moieties in the γ chains of adjacent, aligned fibrin molecules. Solution studies with small labeled peptide inhibitors (such as GPRK) that mimic the N terminus of the fibrin α chain have localized the polymerization surface to residues 337–379 of the fibrinogen γ chain [25], specifically the region around residue Tyr363 [26]. In the rFbgγC30 structure, Tyr363 is a partially exposed residue (Fig. 2a) that, together with the calcium-binding loop, lines a deep pocket called the polymerization pocket or PP. The location of this binding pocket has been confirmed in studies of the structure of the rFbgγC30 co-crystallized with the Gly–Pro–Arg–Pro amide (KPP *et al.*, unpublished results), showing that it occupies a space between the loops of domain P (Fig. 2a). It is significant that a portion of the PP surface is formed by residues that are conserved among the fibrinogen γ chain sequences, but diverge when the complete family of aligned sequences is considered (Fig. 4). Furthermore, 13 residues (between residues 143 and 411 of the fibrinogen γ chain) have been identified as sites of amino acid substitutions, deletions or insertions that lead to calcium-binding defects and/or impaired polymerization. Three of these mutation sites (Gln329 [27], Asp330 [28], and Arg375 [29]) involve sidechains that are partially exposed and line the surface of the PP pocket.

#### Variable surface loops

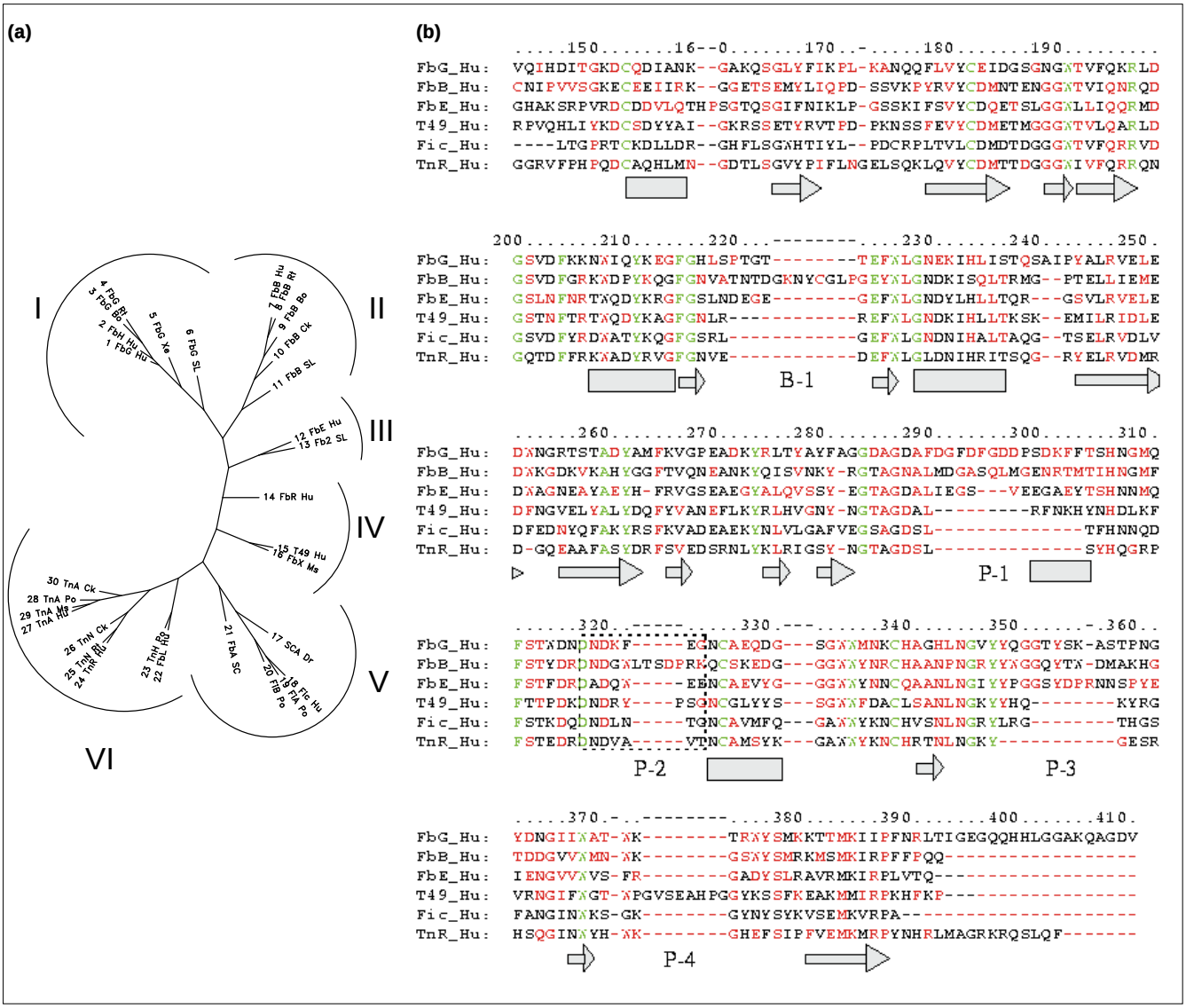
A structure-guided alignment of 30 sequences of the C-terminal portions of fibrinogen γ and β chains and related proteins was performed to investigate the degree of structural similarity to be expected in this family of protein regions. This alignment was used to generate the dendrogram in Figure 4a. Representative sequences for each major group of Figure 4a are aligned in Figure 4b. The protein sequences share from 34% to 99% sequence identity with the human fibrinogen γ chain and are thus expected to have the same protein fold. Most of the 24 residues that are absolutely identical among the 30 sequences have sidechains that are buried in rFbgγC30, or are disulfide cysteines or glycines, and are clearly structurally important. In contrast, the 133 residues that are conserved among the γ-chain sequences are distributed throughout the interior and on the surface of the structure, and are expected to be important both structurally and functionally.

To test the hypothesis that the C-terminal regions of the β and γ chains share a similar fold, the experimental γ chain coordinates were combined with sequence alignment information to construct a homology model of the C terminus of the human β chain (which is 42% identical to rFbgγC30). Both the experimental rFbgγC30 structure and the modeled β-chain structure were analyzed using a residue environment profile method and were found to have scores higher than the average for properly folded proteins of their size [30]. There are three sites of notable structural difference between the C-terminal fragments of the human fibrinogen γ and β chains. The first is an eight-residue insertion in the B-1 loop of the β chain, between two β strands in the B domain (Figs 2a, 4b). The inserted sequence in the fibrinogen β chain contains a cysteine residue which forms a disulfide bond with the N terminus of the A domain. In the rFbgγC30 structure, loop B-1 is solvent-exposed and close to the N terminus. The loop insertion and additional disulfide bond in the β chain are not expected to cause significant changes in conformation in other parts of the structure.

The second pronounced difference between the γ-chain structure and the β-chain model is found in the P-1 loop. In the γ chain of fibrinogen this region contains five conserved aspartic acid residues and presents a negatively-charged surface, whereas the corresponding loop in the fibrinogen β chain is the same length but lacks this acidic character.

The third difference is the insertion of five residues in the P-2 loop of the β chain of fibrinogen, which is the loop that contains the calcium binding site in the γ chain. This insertion would be expected to alter the conformation of the P-2 loop dramatically, since one end is tethered to the rest of the domain by a disulfide bond. The altered conformation would explain the lack of calcium binding in the fibrinogen β chain despite the fact that the calcium ligands, Asp318 and Asp320, are conserved. Differences in the sequence of loop P-1 and the insertion in loop P-2 alter the structural and electrostatic properties of the P domain in the β-chain model. In addition, most of the residues lining the PP pocket in rFbgγC30 are different in the β-chain sequence. Taken together, the differences in structure and sequence in the P domain are consistent with the γ chain being more important in the initial stages

**Figure 4**




of fibrin polymerization than the β chain. These changes would also account for the absence of a calcium-binding site in the β chain, despite a protein fold that is probably conserved in the β chain.

The structure-guided sequence alignment shown in Figure 4b also directs attention to two loops in the P domain that are very different from the corresponding regions in the non-fibrinogen sequences. Loops P-1 and P-3 have similar lengths in the γ and β chains of fibrinogen but are dramatically truncated by three to fourteen residues in the non-fibrinogen sequences. The shortening of the P-1 and P-3 loops is expected to result in the absence of the P-domain pocket in these proteins. Another significant alteration occurs in the T lymphocyte proteins (in Group IV), which have deletions of nine residues in the P-1 loop and insertion of eight residues into the P-4 loop. This P-4 loop precedes the C-terminal β strand of the B domain and is coupled with a truncation of the P-3 loop. One possible consequence of this eight-residue insertion would be to compensate, in part, for the P-1 and P-3 deletions by providing a new, extended surface for the P domain. In rFbgγC30, the P domain contains many of the structural determinants of the protein's function. Sequence alignments show that this domain varies the most among the proteins expected to share the same fold. The pronounced structural variability in this functionally important domain reflects the diversity of function in this family, members of which bind to proteins or to saccharides, or whose ligands, if any, are yet to be identified.

**Figure 4 continued**

Dendrogram and representative sequences from the structure-guided sequence alignment of fibrinogen-related proteins. (a) Dendrogram of the aligned sequences. Each protein was given a number; the first three alphabetic characters of each name define the type of protein fragment and the last two describe the species. The protein sequences are divided into six groups as follows: group I, fibrin γ chains, proteins 1–6; group II, fibrin β chains, proteins 7–11; group III, extended α chains, proteins 12–13; group IV, fibrin related sequences, proteins 14–16; group V, ficolins and invertebrate proteins, proteins 17–21; group VI, tenascins and homologs, proteins 22–30. In this dendrogram, the branch lengths are natural logarithms of the steps between nodes (see Materials and methods). (b) Sequence alignment of a representative human protein from each of the major groups in Figure 4a. The order of the six proteins is the same as the group label of Figure 4a. The first row of numbers are the sequence positions of rFbgγC30 in fibrinogen. Residues or gap positions identical within the group represented by the sequence are printed in red. Those residues identical in all 30 proteins are printed in green. Secondary structure is indicated by arrows (β strands) or rectangles (α helices). Secondary structure assignment has only been made up to residue 392. The loops labeled B-1 and P-1 to P-4 are discussed in the text. The broken box surrounding residues 318–324 delineates the calcium-binding loop. Gaps occurring in all proteins indicate one or more of the proteins is longer than the representative sequences at the gapped location.
The proteins included in the sequence alignment and the dendrogram together with the reference, SWISS-PROT or PIR identifier are as follows: 1 FbG Hu, P02679, human (*Homo sapiens*) fibrinogen γ A chain, residues 143–411; 2 FbH Hu, P04469, human fibrinogen γ B (γ′) chain, residues 143–427; 3 FbG Bo, P12799, bovine (*Bos taurus*) fibrinogen γ B chain (γ′), residues 143–420; 4 FbG Rt, P02680, rat (*Rattus norvegicus*) fibrinogen γ A and B chain, residues 144–420; 5 FbG Xe, P17634, African clawed frog (*Xenopus laevis*) fibrinogen γ chain, residues 141–413; 6 FbG SL, P04115, sea lamprey (*Petromyzon marinus*) fibrinogen γ chain, residues 144–408; 7 FbB Hu, P02675, human fibrinogen β chain, residues 201–461; 8 FbB Rt, P14480, rat fibrinogen β chain (fragments), residues 58–310; 9 FbB Bo, P02676, bovine fibrinogen β chain, residues 187–447; 10 FbB Ck, Q02020, chicken (*Gallus gallus*) fibrinogen β chain (fragment), residues 205–463; 11 FbB SL, P02678, sea lamprey fibrinogen β chain (fragment), residues 220–477; 12 FbE Hu, B44234, human α chain, extended form (α E) residues 18–262; 13 Fb2 SL, P33573, sea lamprey fibrinogen α-2 chain, residues 371–617; 14 FbR Hu, JN0596, human fibrinogen-related protein HFREP-1, residues 73–312; 15 T49 Hu, S47273, human pT49 fibrinogen-like T-lymphocyte protein, residues 203–439; 16 FbX Ms, P12804, mouse (*Mus musculus*) cytotoxic T-lymphocyte specific protein, residues 177–413; 17 SCA Dr, P21520, *Drosophila melanogaster* scabrous protein, residues 488–689; 18 Fic Hu, human ficolin, [58], residues 99–313; 19 FiA Po, A47172, porcine (*Sus scrofa*) ficolin α, residues 109–323; 20 FiB Po, B47172, porcine ficolin β, residues 112–326; 21 FbA SC, P19477, sea cucumber (*Parastichopus parvimensis*) fibrinogen-like protein A (FREP-A), residues 36–258; 22 FbL Hu, P22105, human XB gene overlapping P450c21B, residues 596–820; 23 TnH Po, S28170, porcine tenascin homolog (fragment), residues 1–220; 24 TnR Hu, human tenascin-R, janusin, J1-160/180, [59] residues 1128–1358; 25 TnN Rt, S32023, rat neural recognition molecule J1-160/180, residues 1126–1356; 26 TnN Ck, JH0675, chicken restrictin, neural extracellular matrix, residues 1123–1353; 27 TnA Hu, P24821, human tenascin, residues 1953-2177; 28 TnA Po, S19694, porcine tenascin precursor, residues 1521–1746; 29 TnA Ms, JQ1322, mouse tenascin precursor, residues 1794–2019; 30 TnA Ck, P10039, chicken tenascin precursor, residues 1583–1808. Further information on this figure is available as Supplementary material (published with this paper on the Internet).

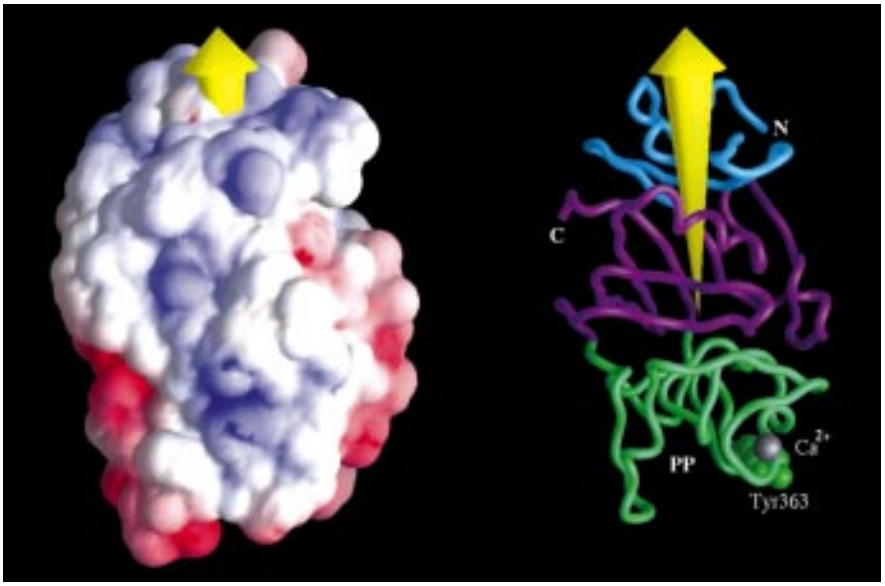
### Molecular dipole

The loops in the P domain are important structural features, since they provide the calcium-binding site and form the polymerization pocket. The calculated surface electrostatic potential of the rFbgγC30 structure shows that these loops also provide the largest areas of negatively charged surface (Fig. 5). The concentration of acidic residues in the P domain results in a large molecular dipole moment (477.7 Debye for the $P2_1$ structure) along the long axis of the fragment. A similar calculation with the homology model of the fibrinogen β chain fragment gives a much smaller molecular dipole moment (137.1 Debye) at an angle of ~70° to the long axis of the structure. Thus the dipole moment of the γ chain is expected to dominate at the distal nodules of the fibrinogen molecule. Electron microscopy has shown that fibrin polymerization begins with the formation of double-stranded protofibrils, in which adjacent strands are aligned in a half-staggered, end-to-middle domain pairing. It is likely that the long-range, regular arrangement of fibrils is due not only to the specific, but limited interaction between the γ chain polymerization pocket and the N terminus of the α chain, but also to the substantial dipole moment of the rFbgγC30 fragment.

### Crosslinking sites

There is some controversy as to whether factor XIIIa catalyzes the formation of longitudinal crosslinks between fibrin γ chains aligned end to end in the same fibril [31] or transverse crosslinks between γ chains in adjacent fibrils [32] (Fig. 6). Based on the location of various known positions in the fibrin molecule, the 225 Å fibrin repeat distance observed in electron-microscopy studies, and the dimensions of the rFbgγC30 structure, it is possible to calculate the distance for a transverse crosslink (Fig. 6) with the following assumptions: first, the disulfide bond marking the twofold-symmetry axis in fibrin is formed between Cys28 on each of the two α chains in fibrin. Second, the N-terminal Gly17, and several additional residues of the α chain of one fibrin monomer fit into the polymerization pocket of an adjacent fibrin monomer, while an additional nine residues from the α chain lie outside the pocket. Third, the linear distance from the binding pocket to residue 392 in the γ chain is approximately 35 Å. Residue 392 is the last one with a fixed position in all three crystal structures. Fourth, ten more residues will extend from residue 392 to the midpoint of the reciprocal factor XIII crosslinks. Accordingly, the distance for 19 residues along with the 35 Å mentioned

**Figure 5**




Molecular dipole moment of the rFbgγC30 structure. (a) Accessible surface with positive and negative electrostatic potential colored blue and red, respectively. The head of the yellow arrow, representing the pronounced molecular dipole moment, is seen at the top of the figure. (b) Coil drawing of the structure in the same orientation as (a), with the dipole moment drawn as a large yellow arrow. The three domains are colored separately and the calcium ion is shown as a gray sphere. The N and C termini of the chain, as well as the polymerization pocket (PP), are also labeled. Figure was produced using GRASP [60].

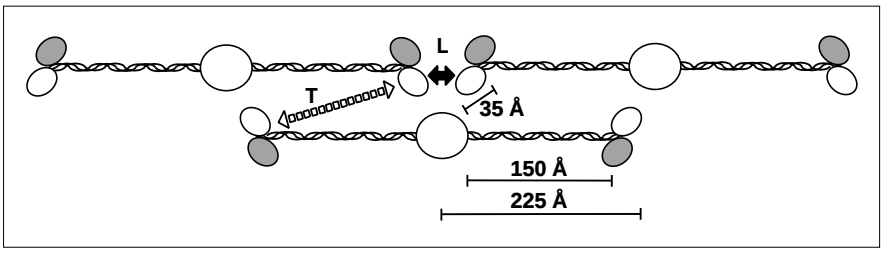
above is needed to reach halfway between α chain globular domains (Fig. 6). Thus the equation for the distance becomes $19x + 35 = 225/2$, where x is the linear distance that must be covered by each residue and 225 Å is the half-length of the fibrin monomer. The value of x is approximately 4 Å, a number too large to span the Cα–Cα length of a peptide (3.8 Å). As predicted in Figure 6, the distance indicated by the label 'T' is longer by 7.6 Å than available in peptides plus the distance through the rFbgγC30 fragment. With the more appropriate figure of 3.5 Å as the mean residue length in an extended peptide, the discrepancy increases to 19 Å. Although the calculation is rough, the requirements for maximal extension of the polypeptide chain and colinearity of the three segments of the line allow us to conclude that transverse crosslinking of the α chains in linear fibrin fibrils is unlikely.

### Flexible C terminus

Fibrin polymerization yields a soft blood clot which is rendered mechanically stable and far more resistant to proteolysis after covalent crosslinking by factor XIIIa. The C terminus of the γ chain contains the Gln398/Gln399 amine acceptor and Lys406 donor residues that are crosslinked [9], as well as the platelet receptor binding site [10]. In the rFbgγC30 $P2_1$ structure, amino acids beyond Leu392 are not visible in any of the electron density maps. HPLC analysis of a cyanogen bromide digest of rFbgγC30 (similar to that from which the $P2_1$ crystals were grown) showed the presence of three species that differed at the C terminus. Mass spectroscopic analysis revealed: one major component ending at His400 and minor components ending at Ala408 and Gly409, a full length protein ending at Val411, and a fragment ending at His401 (data not shown). Since mass spectroscopy of the $P2_1$ crystals themselves failed, it is not known whether the missing residues Thr393–Val411 are invisible due to proteolysis or because they are disordered.

To address this issue, we purified a new batch of protein that was stabilized throughout the purification by the presence of protease inhibitors (N-ethyl maleimide (NEM), phenylmethylsulfonyl fluoride (PMSF), aprotinin and pepstatin A). The crystals grew in group $P2_12_12$ under

**Figure 6**




Schematic representation of the longitudinal (L) versus transverse (T) end-to-end γ–γ crosslinking between adjacent fibrin molecules.

**Figure 7**

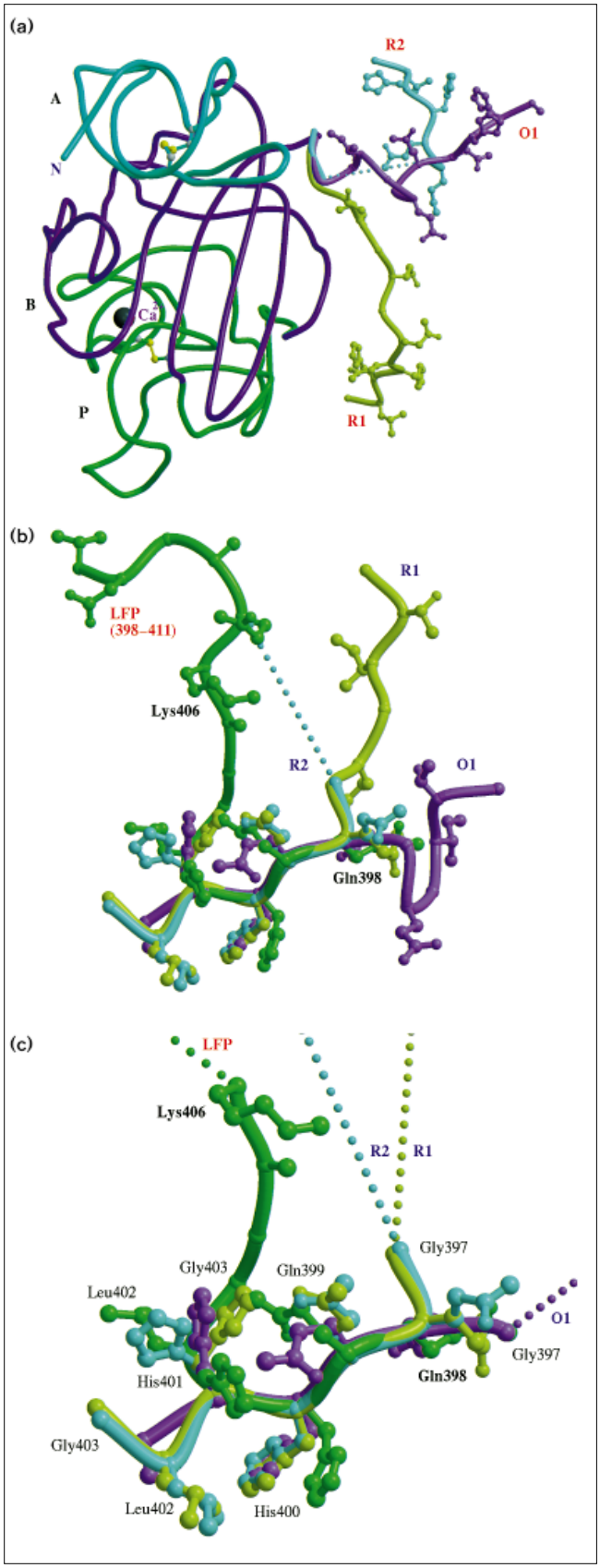



**Figure 7 continued**

Superpositions of extended C-terminal residues. (a) Coil drawing of the three superimposed rFbgγC30 structures. The globular portion of the protein, consisting of residues 144–392, is colored and labeled by domain (A in cyan, B in purple, and P in green); the calcium ion is shown as a gray sphere and labeled, along with the N terminus of the chain. The C-terminal regions of molecules 1 and 2 in the R3 crystal (R1 light green and R2 pale blue), and of the $P2_12_12$ structure (01, lavender) are labeled in red and drawn as peptide coils with superimposed sidechain groups in ball-and-stick form. Residues 393 to 396 of molecule 2 in the R3 crystal could not be traced in the electron-density maps; this missing segment is shown as a string of small pale blue spheres. (b) The C-terminal regions of the rFbgγC30 structures are superimposed, along with residues 398 to 411 from a lysozyme fusion protein X-ray structure (PDB identifier 1LSG). Cα atoms for residues 398–401 were used to generate the superposition. The three rFbgγC30 structures are colored as in (a) and labeled at their N-termini; the peptide from the lysozyme fusion protein is colored dark green and is labeled at its C terminus. (c) Detailed view of the four superimposed C-terminal regions shown in (b), with all common sidechain groups labeled. The paths of the termini are shown as strings of spheres which are in the same color as the corresponding region. As in (b), the N termini of the three rFbgγC30 structures are labeled in blue and the C terminus of the lysozyme fusion structure is labeled in red. Figure was drawn using MOLSCRIPT [56].

crystallization conditions identical to those that produced $P2_1$ crystals in the absence of inhibitors (see Materials and methods). In the R3 and $P2_12_12$ crystal structures, the last visible residues are Gly403 and Leu402, respectively. Mass spectroscopic analysis of the protein solution with protease inhibitors present (NEM and PMSF) and of the data collection crystals indicated the presence of components ending at Gly403 and at Ala405 (data not shown).

The C-terminal regions (beyond residue Leu392) in the two crystallographically-independent molecules of the R3 crystal form and the C-terminal region in the $P2_12_12$ structure are found in three different conformations. These differences appear to be due to the different local crystal packing environments. All of the residues observed beyond Leu392 in the three conformations form intermolecular contacts and/or hydrogen bonds. Residues Thr393– Glu396 are visible in electron-density maps only for molecule 1 in the R3 space group; the corresponding region in molecule 2 is is not visible in the maps. The two independent R3 molecules and that in the $P2_12_12$ space group have similar intermolecular mainchain hydrogen bonding. This may be significant because in each case the hydrogen bonding corresponds to that in β strands. It is tempting to speculate that the tendency of the Gln399–His402 segment to form β strands may be important in the alignment of the C termini during the factor XIIIa covalent crosslinking process.

Despite some similarity in the intermolecular hydrogen-bonding patterns involving the C-terminal regions in the R3 and $P2_12_12$ crystal forms, the three peptides are found in dramatically different positions and orientations relative

to the globular three-domain body of the protein (Fig. 7a). Different crystal packing interactions account for these large differences. When only the extended C-terminal regions are superimposed (Figs 7b, c), another view of the conformational differences among the peptides is provided. Also included in the comparison is the structure of the Gln398–Val411 region, as determined in a crystallographic study of the peptide fused to the C terminus of chicken egg-white lysozyme (PDB identifier 1LSG) [33]. The positions of the four common Cα atoms (from Gln398 to His401) used in the superposition agree well, but the mainchain of the peptides on either side of this short segment deviate substantially. On a local level, the sidechain conformations also differ significantly from each other (Fig. 7c). One factor XIIIa crosslinking site, Gln398, is present in all four structures, and its sidechain is found in four different conformations. The other crosslinking site, Lys406, is only observed in the lysozyme fusion protein; however, its position and orientation relative to the Gln398 residue are not compatible with the peptide conformations observed in the R3 and $P2_12_12$ crystal structures, because their C-terminal residues (Leu402 or Gly403) are in very different positions.

The ability of the γ-chain C-terminal residues to adopt different conformations due to differing intermolecular interactions is consistent with other structural results. Solution NMR studies at pH 3 of a Lys385–Val411 peptide indicated that residues Asn390–Leu402 adopt a multiple-turn, or helix-like structure, whereas residues 403–411 were structurally nondescript [34]. Other solution NMR studies have been conducted on the Leu392– Val411 peptide, which binds to monoclonal antibody 4A5 [35]. These experiments, carried out at pH 5.2, revealed that the structural features necessary for binding of the peptide to the platelet receptor were considerably different from those necessary for antibody binding [35]. This implies that the conformation of the peptide must be different in these two interactions. Finally, an NMR study of the peptide His400–Val411 interacting with the integrin GPIIb/IIIa platelet receptor used distance geometry calculations based on transfer NOE measurements to demonstrate that the peptide was helical when bound to the platelet receptor [36].

The multiple conformations observed for the C-terminal region Thr393–Val411 provide firm evidence that this region is flexible. The C-terminal region is highly susceptible to proteolysis (HCFC, unpublished data) and indeed, mass spectrometry of some of the solutions from which the crystals were grown and two of the crystals themselves indicated that these samples contained protein that was proteolyzed to some degree at the C terminus. Because of this, we cannot fully rule out the possibility that the full-length protein may provide additional stabilizing interactions that 'anchor' this region in a single conformation. Efforts are continuing to produce and crystallize intact rFbgγC30. Our working hypothesis at present is that the flexibility in the C-terminal region is physiologically important, allowing the protein to adapt to crosslinking by factor XIIIa and platelet-receptor binding by adopting different conformations.

#### Implications of the rFbgγC30 structure

Historically, understanding of the process of fibrin polymerization and factor XIIIa-catalyzed crosslinking has been limited by a lack of detailed structural information. Recent crystallographic studies of factor XIIIa have shown that the active site of the enzyme is buried in both the zymogen (inactive) and in the thrombin-cleaved forms [37,38]. It is clear, however, that a conformational change in the enzyme must occur to make the active site accessible to its substrate. The fibrin substrate is itself a cofactor in factor XIII activation, as indicated by solution studies [39]. The conformational flexibility in the crosslinking site of the γ chain of fibrinogen, as indicated in this X-ray crystallographic study and other published NMR solution studies, suggests that conformational changes occur (both in the substrate and in the factor XIIIa enzyme) during the crosslinking process. The pronounced molecular dipole described in this work suggests an orientation mechanism for fibrin alignment during the polymerization process. Finally, the elucidation of the structural details of the calcium-binding site and the polymerization pocket provide valuable information for the design of high affinity fibrin ligands targeted at the γ chain of fibrinogen.

## Biological implications

Vascular damage leads to platelet plug formation and the generation of an insoluble fibrin clot at the site of injury. Fibrinogen is critical in these reactions, which are responsible for reducing, then arresting the loss of blood. Initially, fibrinogen forms a bridge between activated platelets as the platelet plug is being generated. This is followed by fibrin generation, polymerization and crosslinking to yield a more stable platelet plug.

The C-terminal portion of the γ chain of fibrinogen has a significant role in these reactions. It contains the region that binds to activated platelets, the polymerization pocket, the crosslinking site and a calcium-binding site. The three-dimensional structure of a C-terminal fragment of the γ chain of human fibrinogen (rFbgγC30) has made it possible to identify and locate the polymerization pocket, as well as the single calcium-binding site in the molecule. Specific amino acid residues that line the polymerization pocket include Tyr363, which was identified earlier as part of the polymerization site [26]. Seven oxygen atoms coordinate the calcium ion. These atoms are provided by specific aspartic acid residues and mainchain atoms of phenylalanine and glycine, which form the short Asp318–Gly324 calcium-binding loop, in addition to two water molecules. These data clearly establish the requirement for calcium in fibrin formation and fibrinogen stability.

In addition to these well established molecular features, this structure suggests several intriguing possibilities. First, the structure-guided sequence alignment of the C termini of a variety of extracellular proteins places this fragment of the fibrin molecule into a biological context not normally considered for fibrin. The sequence and phylogenetic analysis shows this globular portion of the fibrin molecule is a member of a family of proteins that bind diverse ligands such as proteins and saccharides. Second, the finding of an impressive molecular dipole moment for rFbgγC30, substantially larger than that calculated for the β-chain model, suggests the possibility of electrostatic steering as a significant component of the fibrin polymerization mechanism. Third, the question of whether factor XIIIa crosslinking reactions of fibrin are transverse or longitudinal can now be addressed, and we find the transverse crosslink unlikely. Finally, two of the crystal forms examined possess three different conformations for the C-terminal residues beyond residue 392. This information, taken together with a diversity of conformations observed by others both in solution and in crystals, suggests that conformational flexibility of the C terminus of the γ chains is functionally important in its biological context.

## Materials and methods

### *Preparation of rFbgγC30*

A fragment from the C terminus of the γ chain of human fibrinogen encoding residues Val143–Val411 was expressed in the methylotrophic yeast *Pichia pastoris* (strain GS115) according to the manufacturer's protocol (Invitrogen, San Diego, CA). The vector, pPIC9K, was kindly provided by Dr Michael Romanos (Wellcome Research Laboratories). The recombinant protein (rFbgγC30) was purified to homogeneity by ammonium sulfate precipitation, followed by cation-exchange chromatography. rFbgγC30 inhibited the polymerization of desAfibrin in batroxobin-initiated fibrinogen clotting assays (data not shown), indicating that the polymerization site is biologically functional within this fragment. Moreover, the fragment was readily crosslinked by recombinant human factor XIIIa, a gift from Dr Paul Bishop (ZymoGenetics Inc., Seattle, WA) (data not shown).

### *Crystallization*

Monoclinic crystals of rFbgγC30 (space group $P2_1$, unit cell parameters a = 37.53 Å, b = 68.42 Å, c = 47.67 Å and β = 105.36°) were grown at room temperature by sitting drop vapor diffusion employing two different conditions: 20% PEG3350 and 70 mM $CaCl_2$ in 0.1 M Tris buffered at pH 8.0; and 18% PEG8000 and 70 mM $CaCl_2$ in 0.1 M MES buffered at pH 6.0. When a solution of rFbgγC30 was stabilized in a Tris/NaCl buffer containing the inhibitors NEM and PMSF at pH 7.4 and then dialyzed against a 5 mM MES solution containing NEM and PMSF and buffered at pH 6.0, rhombohedral crystals (space group R3 unit cell parameters a = 164.65 Å and c = 84.88 Å) were grown. The protein solution remaining from the dialysis experiment was used to grow the orthorhombic crystal form (space group $P2_12_12$, unit cell parameters a = 48.70 Å, b = 61.58 Å, and c = 83.25 Å) by room temperature sitting drop vapor diffusion under the same PEG8000 conditions which yielded the original $P2_1$ crystals.

### *Data collection and preparation of heavy atom derivatives*

Since the $P2_1$ crystals were obtained first, they were used in the initial multiple isomorphous replacement (MIR)structure solution. $K_2PtCl_4$ and $UO_2(NO_3)_2$ derivatives were obtained by soaking pH 6.0 crystals in artificial mother liquor with 5 mM of the heavy atom compound added for one day. A $TbCl_3$ derivative was obtained by first soaking in artificial mother liquor containing no added calcium for two days, and then soaking in the presence of 5 mM $TbCl_3$ at pH 8.0 for two days. All diffraction data were collected on either of two systems, one a Siemens X1000 area detector equipped with a graphite monochromator, and the other a Rigaku Raxis-II imaging plate detector with focusing mirrors; each apparatus was mounted on a Rigaku RU200 rotating anode X-ray source supplying CuKα radiation. Siemens data were processed using XENGEN [40] while Raxis data were processed using either Raxis software or DENZO [41] (Table 1).

### *Phase determination and model building*

The $P2_1$ derivative data sets were scaled to native data with KBRAN[l (S Mande and W Hol, unpublished data). Heavy atom sites for each derivative were located by analysis of isomorphous difference Patterson maps, and confirmed by difference Fourier syntheses. Programs in the CCP4 suite [42] were used for heavy atom phase calculation and refinement (MLPHARE) (Table 1), density modification (DM), and map calculation. Model building was carried out with O [43]. An initial polyalanine model was built into 10.0–2.5 Å resolution MIR and density-modified MIR maps which resulted from implementation of the solvent flattening, histogram matching, application of Sayre's equation, and skeletonization protocols in the CCP4 program DM. Phases calculated from the polyalanine model were combined with MIR phases using SIGMAA [42,44] to yield improved electron-density maps into which a complete molecular model could be built.

### *Crystallographic refinement*

All refinement calculations were performed using energy minimization and simulated annealing features in X-PLOR [45]. The starting R factor for the $P2_1$ MIR model of 41.3% ($R_{free}$ = 43.0%) against data from 10.0 to 2.5 Å resolution dropped to 31.8% ($R_{free}$ = 37.9%) in one round of energy minimization. Further refinement of the $P2_1$ model was carried out by alternating between cycles of manual refitting in O with X-PLOR energy minimization and simulated annealing calculations. In the final stages of refinement, the resolution of the diffraction data included in the calculations was extended to 10.0–2.1 Å (Table 2).

### *Molecular replacement*

The R3 and $P2_12_12$ structures were solved by molecular replacement calculations with the refined $P2_1$ coordinates and the program AMoRe [42]. Using 8.0–3.0 Å resolution data, the two independent molecules in the R3 crystal form, and the single molecule in the $P2_12_12$ structure, could be easily located. For the R3 structure, the molecular replacement solution gave a correlation coefficient between calculated and observed Patterson vectors and R factor of 77.5% and 28.9% respectively, while the corresponding values for the $P2_12_12$ structure were 77.6% and 30.1%. Initial refinement of these two crystal forms was carried out using 10.0–2.5 Å resolution data. In the case of the $P2_12_12$ crystal form, final refinement calculations included data from 10.0–2.1 Å resolution. For all three crystal forms, a random 10% of the diffraction data were set aside for the $R_{free}$ calculation in order to monitor the course of refinement (Table 2).

Calcium ions and solvent molecules were located in peaks in difference Fourier maps before being included in the refinement. In the three crystals, the same single calcium-binding site was found in each molecule and its location corresponded to the highest peak of residual electron density, which was greater than 10σ in each case (σ represents the root mean square electron density for the unit cell). During the final cycles of X-PLOR refinement, all diffraction data from 10.0 Å to the high resolution limit were included.

### *Structure verification*

Three different profile methods, 3D1D [30], PROSAII [46], and ERRAT [47], all confirmed the compatibility of the rFbgγC30 crystal structure with its amino acid sequence. The program PROCHECK [48] was used to assess the geometric quality of the three crystal structures and determined that all were of above average stereochemical quality when

**Table 1**

**Data collection and multiple isomorphous replacement phasing.**

| Crystal | $P2_1$ (native) | $K_2PtCl_4$ | $TbCl_3$ | $UO_2$ | R3 (native) | $P2_12_12$ (native) |
|---|---|---|---|---|---|---|
| a (Å) | 37.53 | 37.12 | 37.24 | 37.29 | 164.65 | 48.70 |
| b (Å) | 68.42 | 67.70 | 67.99 | 68.02 | | 61.58 |
| c (Å) | 47.67 | 46.90 | 47.18 | 47.14 | 84.88 | 83.25 |
| β (°) | 105.36 | 105.01 | 105.36 | 105.31 | | |
| $V_M$ (Å$^3$/Da) | 1.97 | | | | 3.69 | 2.18 |
| Detector | Siemens | Siemens | Raxis | Siemens | Raxis | Raxis |
| Resolution (Å) | 33.0–2.0 | 33.9–2.0 | 37.8–2.1 | 31.8–2.0 | 33.3–2.5 | 50.0–2.0 |
| $R_{sym}$(I) (%)* | 3.33 | 3.19 | 5.87 | 5.12 | 8.3 | 6.1 |
| Completeness (%) | 82.7 | 85.3 | 89.8 | 76.4 | 89.2 | 94.6 |
| Number of sites | | 1 | 1 | 4 | | |
| Resolution (Å) | | 10.0–2.5 | 10.0–2.5 | 10.0–2.5 | | |
| $R_{iso}$ (F) (%)† | | 14.6 | 9.6 | 14.2 | | |
| $R_{cullis}$ (centric/acentric)‡ | | 0.66/0.79 | 0.76/0.86 | 0.59/0.51 | | |
| Phasing power (c/a)§ | | 1.05/1.32 | 0.75/0.95 | 1.77/2.91 | | |

*$R_{sym} = \Sigma|I_h - \langle I_h \rangle| / \Sigma I_h$, †$R_{iso} = \Sigma||F_{PH}| - |F_P|| / \Sigma|F_{PH}|$, ‡$R_{cullis} = \Sigma|(F_P + F_{Hcalc})| - |F_{PH}| / \Sigma||F_{PH}| - |F_H||$ and §Phasing power = $\Sigma|F_{Hcalc}| / \Sigma|(F_P + F_{Hcalc}) - F_{PH}|$. Overall figures of merit for centric, acentric and all data are 0.87, 0.61 and 0.63, respectively.

compared with the reference database of structures deposited in the Protein Data Bank.

### *Sequence alignment and homology modeling*

The UWGCG package of programs [49] was used for extraction of proteins from the SWISS-PROT and PIR databases. Family alignments were performed by an updated [50] version of the algorithm described earlier [51]. The PROFILESEARCH method [52] was used to search both databases for further entries, which were then aligned with the family; some newer sequences were entered manually. After the alignment of the sequences, the $P2_1$ model was examined for compatibility with the alignment and the gaps and insertions moved, if necessary, to positions compatible with the protein structure. In general this involved moving a residue or two. The primary criterion for gap placement was surface accessibility and loops; gaps within secondary structure elements were avoided whenever possible. Additionally, insertion and deletion positions (gaps) were chosen to minimize any structural changes, particularly in the core of the protein. The homology model of the β-chain fragment was constructed from the sequence alignment using the BIOSYM software [53].

The phylogenetic analysis of the aligned sequences was performed using maximum parsimony with the PHYLIP package of programs [54]. The best trees were found using PROTPARS. The only uncertainty in the trees concerned the relations among proteins 27, 28, and 29 of Figure 4a. CONSENSE was used to determine the consensus tree among the three possibilities. Following CONSENSE, the ANCSTR program [55] was used to determine leg lengths in the tree. For plotting,

**Table 2**

**Refinement statistics.**

| Crystal | $P2_1$ (native) | R3 (native) | $P2_12_12$ (native) |
|---|---|---|---|
| Resolution (Å) | 10.0–2.1 | 10.0–2.5 | 10.0–2.1 |
| No. of reflections | 11 854 | 26 046 | 15 050 |
| Completeness (%) | 87.8 | 89.1 | 98.9 |
| No. of protein atoms | 1995 | 4132 | 2070 |
| No. of calcium ions | 1 | 2 | 1 |
| No. of solvent | 177 | 202 | 151 |
| $R_{cryst}$ (%)* | 15.5 | 16.8 | 16.7 |
| $R_{free}$ (%)* | 22.7 | 24.4 | 22.6 |
| $R_{ref}$ (%)* | 15.9 | 17.1 | 17.0 |
| Rms deviations | | | |
| bonds (Å) | 0.009 | 0.010 | 0.009 |
| angles (°) | 1.656 | 1.737 | 1.699 |
| dihedral angles (°) | 24.495 | 24.274 | 23.948 |
| improper torsion angles (°) | 1.244 | 1.306 | 1.291 |
| Protein mean B/max B (Å$^2$) | 11.8/65.0 | 21.5/88.5 | 15.1/72.5 |
| Calcium B (Å$^2$) | 14.9 | 24.9, 32.4 | 11.1 |
| Solvent mean B/max B (Å$^2$) | 31.4/60.7 | 38.6/77.7 | 28.6/60.1 |
| Luzzati rms coordinate error (Å) | 0.18 | 0.24 | 0.19 |

*$R_{cryst}/R_{free}/R_{ref} = \Sigma||F_{Pobs}| - |F_{Pcalc}|| / \Sigma|F_{Pobs}|$, where $R_{free}$ was calculated using a random 10 % of the reflections, $R_{cryst}$ was calculated using the remaining 90 % of the reflections, and $R_{ref}$ was calculated after the final cycles of refinement in which all the data were used.

the programs FITCH, RETREE, and DRAWTREE [54] were used. The natural logarithms of the branch lengths were used for plotting Figure 4a. This operation allows the display of all of the sequences without serious overlap of the labels in the figure. Thus the branch lengths in Figure 4a indicate logarithmic distances among the proteins.

### *Accession numbers*

Coordinates for the three crystal forms have been deposited with the Protein Data Bank ($P2_1$, R3, and $P2_12_12$ structures with identifiers 1FIB, 1FIC, and 1FID, respectively).

### *Supplementary material*

The supplementary material contains the 30 sequences which were aligned to construct Figures 4a and 4b. The legend of Figure 4 gives the names of each of the proteins, which are numbered and aligned in the same order as in the Figure. This information is also available at http://www.bmsc.washington.edu/people/teller.

## Note added in proof

An abstract has recently appeared describing crystallography of fragment D and derivatives (Doolittle, R.F., Spraggon, G., Everse, S.J., Veerapandian, L. & Riley, M. (1996). Crystal structure studies on fragment D and double-D from human fibrinogen. *Blood* 88 (suppl. 1), 283).

## Acknowledgements

We thank Focco van den Akker and Professors Elinor Adman and Wim Hol for valuable discussions. We also thank Professor Wim Hol for access to computational resources, Dr Jeff Kowalak for carrying out mass spectroscopic analyses, and Leslie Boba for help in preparation of the manuscript. HCFC was supported by a Research Fellowship from the Heart and Stroke Foundation of Canada. This work was supported by the NIH Research Grants HL50355 and HL16919.

## References

1. Fowler, W.E. & Erickson, H.P. (1979). The trinodular structure of fibrinogen: confirmation by both shadowing and negative stain electron microscopy. *J. Mol. Biol.* 134, 241–249.
2. Fowler, W.E., Fretto, L.J., Erickson, H.P. & McKee, P.A. (1980). Electron microscopy of plasmic fragments of human fibrinogen as related to trinodular structure of the intact molecule. *J. Clin. Invest.* 66, 50–56.
3. Price, T.M., Strong, D.D., Rudee, M.L. & Doolittle, R.F. (1981). Shadowcast electron microscopy of fibrinogen with antibody fragments bound to specific regions. *Proc. Natl. Acad. Sci. USA* 78, 200–204.
4. Mosesson, M.W., Hainfeld, J., Wall, J. & Haschemeyer, R.H. (1981). Identification and mass analysis of human fibrinogen molecules and their domains by scanning transmission electron microscopy. *J. Mol. Biol.* 153, 695–718.
5. Williams, R.C. (1983). Morphology of fibrinogen monomers and of fibrin protofibrils. *Ann. NY Acad. Sci.* 408, 180–193.
6. Mosesson, M.W. (1990). Fibrin polymerization and its regulatory role in hemostasis. *J. Lab. Clin. Med.* 116, 8–17.
7. Doolittle, R.F. & Laudano, A.P. (1980). Synthetic peptide probes and the location of fibrin polymerization sites. *Protides Biol. Fluids* 28, 311–316.
8. Chen, R. & Doolittle, R.F. (1969). Identification of the polypeptide chains involved in the cross-linking of fibrin. *Proc. Natl. Acad. Sci. USA* 63, 420–427.
9. Chen, R. & Doolittle, R.F. (1971). γ–γ cross-linking sites in human and bovine fibrin. *Biochemistry* 10, 4486–4491.
10. Kloczewiak, M., Timmons, S. & Hawiger, J. (1983). Recognition site for the platelet receptor is present on the 15-residue carboxy-terminal fragment of the γ chain of human fibrinogen and is not involved in the fibrin polymerization reaction. *Thromb. Res.* 29, 249–255.
11. Marguerie, G., Chagniel, G. & Suscillon, M. (1977). The binding of calcium to bovine fibrinogen. *Biochim. Biophys. Acta* 490, 94–103.
12. Nieuwenhuizen, W. & Haverkate, F. (1983). Calcium-binding regions in fibrinogen. *Ann. NY Acad. Sci.* 408, 92–96.
13. Boyer, M.H., Shainoff, J.R. & Ratnoff, O.D. (1972). Acceleration of fibrin polymerization by calcium ions. *Blood* 39, 382–387.
14. Haverkate, F. & Timon, G. (1977). Protective effect of calcium in the plasmin degradation of fibrinogen and fibrin fragment D. *Thromb. Res.* 10, 803–812.
15. Ly, B. & Godal, H.C. (1972). Denaturation of fibrinogen, the protective effect of calcium. *Haemostasis.* 1, 204–209.
16. Kleywegt, G.J. & Jones, T.A. (1994). Halloween...masks and bones. In *From First Map to Final Model*. (Bailey, S., Hubbard, R. & Waller, D., eds), pp. 59–66, SERC Daresbury Laboratory, Warrington, UK.
17. Murzin, A.G., Brenner, S.E., Hubbard, T. & Chothia, C.J. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* 247, 536–540.
18. Doolittle, R.F. (1992). A detailed consideration of a principal domain of vertebrate fibrinogen and its relatives. *Protein Sci.* 1, 1563–1577.
19. Koopman, J., Haverkate, F., Briet, E. & Lord, S.T. (1991). A congenitally abnormal fibrinogen (Vlissingen) with a 6-base deletion in the γ-chain gene, causing defective calcium binding and impaired fibrin polymerization. *J. Biol. Chem.* 266, 13456–13461.
20. Varadi, A. & Sheraga, H.A. (1986). Localization of segments essential for polymerization and for calcium binding in the γ chain of human fibrinogen. *Biochemistry* 25, 519–528.
21. Dang, C.V., Ebert, R.F. & Bell, W.R. (1985). Localization of a fibrinogen calcium-binding site between γ-subunit positions 311 and 336 by terbium fluorescence. *J. Biol. Chem.* 260, 9713–9719.
22. Abola, E.E., Bernstein, F.C., Bryant, S.H., Koetzle, T.F. & Went, J. (1987). Protein Data Bank. In *Crystallographic Databases: Information Content, Software Systems, Scientific Applications.* (Allen F.H., Bergerhoff, G. & Sievers, R., eds), pp. 107–132, Data Commission of the International Union of Crystallography, Bonn/Cambridge/Chester.
23. Weis, W.I., Kahn, R., Fourme, R., Drickamer, K. & 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.
24. McPhalen, C.A., Strynadka, N.C. & James, M.N. (1991). Calcium binding sites in proteins: a structural perspective. *Adv. Protein Chem.* 42, 77–144.
25. Shimizu, A., Nagel, G.M. & Doolittle, R.F. (1992). Photoaffinity labeling of the primary fibrin-polymerization site: isolation and characterization of a labeled cyanogen bromide fragment corresponding to γ-chain residues 337–379. *Proc. Natl. Acad. Sci. USA* 89, 2888–2892.
26. Yamazumi, K. & Doolittle, R.F. (1992). Photoaffinity labeling of the primary fibrin polymerization site: localization of the label to γ-chain Tyr363. *Proc. Natl. Acad. Sci. USA* 89, 2893–2896.
27. Miyata, T., Furukawa, K., Iwanaga, S., Takamatsu, J. & Saito, H. (1989). Fibrinogen Nagoya, a replacement of glutamine-329 by arginine in the γ-chain that impairs the polymerization of fibrin monomer. *J. Biochem.* 105, 10–14.
28. Reber, P., *et al.*, & Beck, E.A. (1986). Characterization of fibrinogen Milano I: amino acid exchange 330Asp–Val impairs fibrin polymerization. *Blood* 67, 1751–1756.
29. Yoshida, N., *et al.*, & Asakura, S. (1992). Characterization of an abnormal fibrinogen Osaka V with the replacement of γ-arginine 375 by glycine. *J. Biol. Chem.* 267, 2753–2759.
30. Luthy, R., Bowie, J.U. & Eisenberg, D. (1992). Assessment of protein models with three-dimensional profiles. *Nature* 356, 83–85.
31. Weisel, J.W., Francis, C.W., Nagaswami, C. & Marder, V.J. (1993). Determination of the topology of factor XIIIa-induced fibrin γ-chain cross-links by electron microscopy of ligated fragments. *J. Biol. Chem.* 268, 26618–26624.
32. Mosesson, M.W., Siebenlist, K.R., Hainfeld, J.F. & Wall, J.S. (1995). The covalent structure of factor XIIIa crosslinked fibrinogen fibrils. *J. Struct. Biol.* 115, 88–101.
33. Donahue, J.P., Patel, H., Anderson, W.F. & Hawiger, J. (1994). Three-dimensional structure of the platelet integrin recognition segment of the fibrinogen γ chain obtained by carrier protein-driven crystallization. *Proc. Natl. Acad. Sci. USA* 91, 12178–12182.
34. Mayo, K.H., Bruke, C., Lindon, J.N. & Kloczewiak, M. (1990). $^1$H NMR sequential assignments and secondary structure analysis of human fibrinogen γ-chain C-terminal residues 385–411. *Biochemistry* 29, 3277–3286.
35. Blumenstein, M., Matsueda, G.R., Timmons, R. & Hawiger, J. (1992). A β-turn is present in the 392–411 segment of the human fibrinogen γ-chain. Effects of structural changes in this segment on affinity to antibody 4A5. *Biochemistry* 31, 10692–10698.
36. Mayo, K.H., *et al.*, & Andrade-Gordon, P. (1996). Integrin receptor gpIIB/IIIa bound state conformation of the fibrinogen γ-chain C-terminal peptide 400–411: NMR and transfer NOE studies. *Biochemistry* 35, 4434–4444.
37. Yee, V.C., Pedersen, L.C., Le Trong, I., Bishop, P.D., Stenkamp, R.E. & Teller, D.C. (1994). Three-dimensional structure of a transglutaminase:

human blood coagulation factor XIII. *Proc. Natl. Acad. Sci. USA* **91**, 7296–7300.
38. Yee, V.C., Pedersen, L.C., Bishop, P.D., Stenkamp, R.E. & Teller, D.C. (1995). Structural evidence that the activation peptide is not released upon thrombin cleavage of factor XIII. *Thromb. Res.* **78**, 389–397.
39. Lewis, S.D., Janus, T.J., Lorand, L. & Shafer, J.A. (1985). Regulation of formation of factor XIIIa by its fibrin substrates. *Biochemistry* **24**, 6772–6777.
40. Howard, A.J., Gilliland, G.L., Finzel, B.C., Poulos, T.L., Ohlendorf, D.H. & Salemme, F.R. (1987). The use of an imaging proportional counter in macromolecular crystallography. *J. Appl. Cryst.* **20**, 383–387.
41. Otwinowski, Z. (1988). DENZO: a program for automatic evaluation of film densities. Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT, USA.
42. Collaborative Computational Project No. 4 (1994). The CCP4 Suite: programs for protein crystallography. *Acta Cryst. D* **50**, 760–763.
43. Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron-density maps and the location of errors in these models. *Acta Cryst. A* **47**, 110–119.
44. Read, R.J. (1986). Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Cryst. A* **42**, 140–149.
45. Brünger, A.T. (1992). X-PLOR, Version 3.1. A System for X-ray Crystallography and NMR. Yale University, New Haven, CT, USA.
46. Sippl, M.J. (1993). Recognition of errors in three-dimensional structures of proteins. *Proteins* **17**, 355–362.
47. Colovos, C. & Yeates, T.O. (1993). Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci.* **2**, 1511–1519.
48. Laskowski, R.A., MacArthur, M.W., Moss, D.S. & Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283–291.
49. Program Manual for the Wisconsin Package, Version 8 (1994). Genetics Computer Group, Madison, Wisconsin.
50. Clark, K.L. (1985). Sequence alignments without the use of arbitrary parameters. PhD dissertation, Kansas State University, KS, USA.
51. de Haen, C., Swanson, E. & Teller, D.C. (1976). The evolutionary origin of proinsulin: amino acid sequence homology with the trypsin-related serine proteases detected and evaluated by new statistical methods. *J. Mol. Biol.* **106**, 639–661.
52. Gribskov, M., McLachlan, A.D. & Eisenberg, D. (1987). Profile analysis: detection of distantly related proteins. *Proc. Natl. Acad. Sci. USA* **84**, 4355–4358.
53. Insight II User Guide, Version 2.3.0. (1993). Biosym Technologies, San Diego, CA, USA.
54. Felsenstein, J. (1993). PHYLIP (Phylogeny Inference Package) Version 3.5c. Department of Genetics, University of Washington, Seattle.
55. Fitch, W.M. & Farris, J.S. (1974). Evolutionary trees with minimum nucleotide replacement from amino acid sequences. *J. Mol. Evol.* **3**, 263–278.
56. Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* **24**, 946–950.
57. Merritt, E.A. & Murphy, M.E.P. (1994). Raster3D Version 2.0: a program for photorealistic molecular graphics. *Acta Cryst. D* **50**, 869–873.
58. Matsushita, M., *et al.*, & Mizuochi, T. (1996). A novel human serum lectin with collagen-and fibrinogen-like domains that functions as an opsonin. *J. Biol. Chem.* **271**, 2488–2454.
59. Carnemolla, B., Leprini, A., Borsi, L., Querze, G., Urbini, S. & Zardi, L. (1996). Human tenascin-R. Complete primary structure, pre-mRNA alternative splicing and gene localization on chromosome 1q23-q24. *J. Biol. Chem.* **271**, 8157–8160.
60. Nicholls, A.J. (1993). GRASP Manual. Columbia University, NY, USA.