New structural motifs on the chymotrypsin fold and their potential roles in complement factor B

Hua Jing^{1,2,3}, Yuanyuan Xu⁴, Mike Carson¹,
Dwight Moore¹, Kevin J. Macon^{1,4},
John E. Volanakis^{4,5} and
Sthanam V.L. Narayana^{1,6}

¹Center for Macromolecular Crystallography, School of Optometry,

²Graduate Program in Biophysical Sciences and Pharmaceutical Design Program and ⁴Division of Clinical Immunology and Rheumatology, Department of Medicine, University of Alabama at Birmingham, Birmingham, AL 35294, USA and ⁵Biomedical Sciences Research Center 'A.Fleming', Vari 16672, Greece

³Present address: Center for Blood Research, Harvard Medical School, Boston, MA 02115, USA

⁶Corresponding author
e-mail: narayana@pearl.cmc.uab.edu

Factor B and C2 are two central enzymes for complement activation. They are multidomain serine proteases and require cofactor binding for full expression of proteolytic activities. We present a 2.1 Å crystal structure of the serine protease domain of factor B. It shows a number of structural motifs novel to the chymotrypsin fold, which by sequence homology are probably present in C2 as well. These motifs distribute characteristically on the protein surface. Six loops surround the active site, four of which shape substrate-binding pockets. Three loops next to the oxyanion hole, which typically mediate zymogen activation, are much shorter or absent. Three insertions including the linker to the preceding domain bulge from the side opposite to the active site. The catalytic triad and non-specific substrate-binding site display active conformations, but the oxyanion hole displays a zymogen-like conformation. The bottom of the S1 pocket has a negative charge at residue 226 instead of the typical 189 position. These unique structural features may play different roles in domain–domain interaction, cofactor binding and substrate binding.

Keywords: complement system/domain structure/
factor B/protein–protein interaction/serine protease

Introduction

Serine proteases with a chymotrypsin-like fold (SP) are ubiquitous and involved in a wide range of biological processes. Structural and functional studies on this family of proteases are directed to answering three major questions: what determines their catalytic efficiency, how do they recognize substrates and how is their activity regulated? Extensive pioneering studies on digestive SPs such as chymotrypsin, trypsin and elastase have provided fundamental answers to these questions. The catalytic efficiency is provided by a His57–Asp102–Ser195 triad, located near the oxyanion hole (reviewed in Kraut, 1977).

The substrate-binding specificity and affinity are determined mostly by the geometrical and chemical nature of the substrate-binding pockets, especially the S1 pocket (reviewed in Perona and Craik, 1995; Czapinska and Otlewski, 1999). Blocking these pockets with synthetic peptide analogs or natural protein inhibitors provided insights into the mechanisms of SP inactivation (reviewed in Bode and Huber, 1992). The activation of an SP is attributed in most cases to the zymogen activation where a freely liberated N-terminus induces conformational changes in and around the active site (reviewed in Huber and Bode, 1978; Khan and James, 1998). In some cases, however, SP activation also requires specific protein–protein interactions such as enzyme–cofactor association (Banner *et al.*, 1996; Wang *et al.*, 1998), enzyme–substrate interactions (Volanakis and Narayana, 1996; Jing *et al.*, 1998) and enzyme self-assembly (Pereira *et al.*, 1998). Structural studies of these novel SPs have begun to reveal the diversity of their regulation mechanisms.

The complement system, rich in such regulated SPs, contains eight proteolytic enzymes, all of which belong to the chymotrypsin fold group and cleave specific peptide bonds next to arginine residues (reviewed in Arlaud *et al.*, 1998). Except for factor D, which has a single domain and interacts only with its natural substrate, all others have multiple domains and interact not only with the substrate, but also with certain cofactor proteins. Among these enzymes, factor B (FB) and C2 are homologous, and each supplies a catalytic subunit for the central C3/C5 convertases of complement. They also display structural and functional features distinct from those of the known SPs.

FB and C2 have three domains each: an N-terminal Ba or C2b domain composed of three complement control protein (CCP) modules, followed by a von Willebrand factor type A (vWFA) domain and a C-terminal SP domain. In the presence of Mg²⁺, FB binds cofactor C3b to be activated by factor D, whereas C2 binds cofactor C4b and is activated by C1s (reviewed in Volanakis, 1988). The activation results in the release of their N-terminal domains and the generation of C3b-bound Bb or C4b-bound C2a, respectively, and the consequent protein complexes are named C3 convertases. Hence, the C-terminal SP domains of the catalytic subunits Bb and C2a lack the typical free N-terminus essential for the enzyme activation due to their linkage to the vWFA domains.

The esterolytic activities of FB/Bb and C2/C2a towards their best synthetic substrates are ~10³-fold lower than that of trypsin (Kam *et al.*, 1987), indicating a possible defect in the catalytic apparatus. No natural inhibitor has been identified for either enzyme. Instead, they are regulated by an association with the respective cofactor proteins. On the other hand, this bimolecular assembly

Table I. Statistics on diffraction data and refined structure

	Room temperature data			Low temperature data	
	Native	TMLA	Sm	Native	TMLA
Diffraction data and phasing statistics					
Cell dimensions	51.88, 75.94, 74.04, β = 96.65			51.64, 74.26, 73.26, β = 95.55	
Resolution (Å)	2.3	2.8	2.8	2.1	2.3
Completeness (%)	97.7	96.0	91.0	99.2	97.5
R_{sym} (%)	7.3	6.0	4.1	6.9	8.1
R_{iso} (%)	—	12.2	8.9	—	16.4
Phasing power (acentric/centric)		1.05/0.88			1.33/1.00
tR_{Cullis}		0.84/0.88			0.78/0.70
Figure-of-merit		0.44/0.69			0.30/0.57
Refinement and structure statistics					
Resolution (Å)	20–2.1				
No. of unique reflections	32 125				
R_{cryst}	20.9				
R_{free}	24.3				
No. of protein atoms	4508				
No. of water atoms	517				
Refined residues					
	A: A1k-V4, K9-K220a, Q220h-L250				
	B: S1b-K220a, V220i-L250				
Invisible residues					
	A: W5-E6-H7-R8, N220b-Q220c-K220d-R220e-Q220f-K220g				
	B: A1k-D1j-Pli-D1h-E1g-S1f-Q1e-S1d-L1c, N220b-Q220c-K220d-R220e-Q220f-K220g-Q221h				
R.m.s.d. bond length (Å)	0.009				
R.m.s.d. bond angle (°)	1.43				
Luzzati error (work/test set) (Å)	0.26/0.30				
No. of residues (A/B)	288/282				
Average <i>B</i> -factor for protein (A/B)	31.99/24.04				
Average <i>B</i> -factor for water	37.3				

is unstable and exhibits a half-life of <2 min under physiological conditions (Kerr, 1980; Pangburn and Müller-Eberhard, 1986). The dissociated Bb displays residual hemolytic and proteolytic activity (Fishelson and Müller-Eberhard, 1984), albeit with higher esterolytic activity than FB (Kam *et al.*, 1987). The cofactor interacts with the Mg^{2+} -binding motif on the vWFA domain in Bb/C2a (Horiuchi *et al.*, 1991; Hourcade *et al.*, 1995; Tuckwell *et al.*, 1997) and possibly also with the SP domain (Lambris and Müller-Eberhard, 1984; Sanchez-Corral *et al.*, 1990). Very little is known about the mode of cofactor binding and the position of the binding site(s) relative to the active site.

To understand the regulation mechanism and to facilitate the design of complement inhibitors, we determined the crystal structure of the SP domain of FB (FBSP). The structure reveals many motifs novel to the chymotrypsin fold. Sequence homology with C2SP suggests that similar motifs are likely to be present in C2SP as well, although some surface features might be different. The structure is presented below with an emphasis on the novel motifs. Some suggestions on their functional correlates, such as domain–domain interaction and cofactor binding, are presented.

Results

Overall structure

The FBSP structure was solved by a combination of multiple isomorphous replacement (MIR) and molecular replacement methods (Table I). All residues are in the most favored (85.6%) or additionally allowed (14.2%)

regions of the Ramachandran plot except for one (K144) in the generously allowed region. The completeness of the two molecules in the asymmetric unit is slightly different for the N-terminal linker region and a surface loop (Table I). The rest of the two molecules superimpose well with a root-mean-square deviation of 0.34 Å for 277 matched (<3.8 Å) C_α atoms.

FBSP displays a two-domain β -barrel fold in which two β -sheets, composed of six β -strands each, are surrounded by surface helices and loops (Figure 1A and B). Compared with other SPs, the core β -sheets are similar, but many surface loops and helical regions are strikingly different. Long insertions and deletions are quite distinct in the structure-based sequence alignment with other SPs (Figure 1C). Interestingly, the insertions and deletions are clustered characteristically on the back, the right side and around the active site (refer to the orientation in Figure 1A). The active site also displays certain atypical conformations for some critical residues.

Three insertions on the back

The N-terminal linker region (residues D1h–T11 in chymotrypsinogen numbering), the 125 insertion (T125a–D133) and the C-terminal insertion (Q243–L250) are clustered on the back of the structure (Figure 1B). The N-terminal linker associates with the main body of the SP domain through a disulfide bond between C1 and C122. The N-terminal end of this linker displays a helical conformation in one molecule but is disordered in the other molecule; the C-terminal end displays a flexible loop conformation. The 125 insertion folds into a helix–loop–helix motif where the N-terminal ends of the two helices are constrained by a unique disulfide

bond (C125–C125p). The second helix is present at a similar position in a few other SPs, but the first helix and the middle loop regions are unique (Figure 1C). The C-terminal insertion is a loop whose conformation is stabilized by specific hydrophobic and electrostatic interactions involving its last four residues.

These three regions are tightly associated with each other and with the main body by a network of disulfide bonds, hydrophobic interactions and hydrogen bonds. A nearby short 203 loop (K203–R206) enhances such an association by complementing with salt bridge partners. The combined motif is obviously rigid and displays three distinct positively charged, negatively charged and hydrophobic patches (Figure 1B).

This motif may display a comparable conformation in C2SP, as indicated by similar residue compositions in the four involved regions (Figure 1C). Some differences, however, can be identified for a few exposed residues in the linker region (residues 3–9) and the C-terminal insertion (residues 245–246). Particularly, N5 and N9 in C2SP are potential N-linked glycosylation sites that are probably occupied (see Xu and Volanakis, 1999), suggesting that this segment might be surface exposed in C2/C2a. Mutation of a glycine at position 2 to an arginine was described as the cause of C2 secretion deficiency (Westel *et al.*, 1996), probably due to the disruption of the nearby C1–C122 disulfide bond.

Three deletions on the right side

Three loops that are present in all other SPs, the 16 loop (16–23), the 71 loop (71–81) and the 142 loop (142–155), are nearly absent in FBSP (Figure 1C). The 16 loop is completely absent, while loops 71 and 142 have only eight residues in total as compared with 24–29 residues observed in other SPs. Notably, seven of the eight residues are charged, but most of them do not form salt bridges. Thus, these three deletions create a unique hydrophilic surface to the right of the active site.

Similarly, C2SP has no 16 loop. Its 71 loop is longer than that in FBSP (seven residues versus three), its 142 loop is shorter (two residues versus five) and only two of the nine residues are charged. The 142 loop contains another potential N-linked glycosylation site at N142, which would make the deletion patch also hydrophilic in C2/C2a. These three loops typically mediate zymogen activation in other SPs (see Jing *et al.*, 1999). Their consensus deletion in FBSP and C2SP may be related to their unique regulation mechanism.

Six loops around the active site

The active site is surrounded by loops 35, 60, 96, 172, 186 and 218, among which loops 96, 172 and 218 are exceptionally long (Figure 1). The 35 loop (V35–E40) and 60 loop (T60–S63) present on top of the substrate-binding cleft shape the upper wall of the S1' and S2'

pockets with the side chains of T60, E40 and S41. The long 96 loop (I96–F98) and 172 loop (A172–V175) interdigitate with each other on the left side of the active site. The side chain of Y99, by stacking with that of W215, divides the space between the S2 and S3 pockets (Figure 2B). The S2 pocket is small and the S3 is large, both being hydrophobic in nature, and correlate well with the specificity towards alanine or glycine at the P2 position and leucine at the P3 position of the natural substrates. These two pockets appear to be rigid because of the extensive interactions between loops 96 and 172. This may explain their strict substrate-binding specificity and imply that they are unlikely to undergo any substrate-induced conformational changes. Similar conformations may be expected for the corresponding loops in C2SP, as indicated by their high sequence homology (Figure 1C).

The 186 loop (V186–D187) is located underneath the S1 pocket (Figure 1A). It interacts with a short 163–164c loop through a buried salt bridge between D187 and K163 (Figure 2A) and an aromatic stacking between Y186c and Y161, both of which are unique to FBSP. The 186 loop in C2SP carries somewhat different residues, but the salt bridge between D187 and K163 might be conserved (Figure 1C).

The 218 loop (V218–A224) is in front of the S1 pocket. Three exposed valine residues and a partially buried D218a in the beginning of this loop line the entrance to the S1 pocket (Figure 2B). C220 forms a disulfide bond with C191, which, due to insertions, is shifted by a few angstroms from its typical position (Figure 2A). Following C220, a stretch of approximately six unique charged or hydrophilic residues is invisible in the electron density maps (Table I). The corresponding region in C2SP is longer and has similar residue composition (Figure 1C), and thus may also display a highly flexible conformation. This atypically long and flexible loop, especially in proximity to the unique deletion patch and active site, probably contributes to the cofactor binding as discussed below.

Within the active site

The catalytic triad residues (H57, D102 and S195) and the non-specific substrate-binding site (W215–G216) display typical active conformations (Figure 2A). However, the oxyanion hole exhibits an atypical conformation due to an inward orientation of R192 carbonyl oxygen (Figure 2B). It forms a hydrogen bond with the amide group of S195, thus reducing the typical positive charge expected in the oxyanion hole. Residues 191–194 adopt a single turn 3_{10} -helix conformation, whose C-terminal end imposes additional negative potential to the oxyanion hole (Figure 2A). Such an inactive oxyanion hole conformation is similar to that observed in some SP zymogens (see Jing *et al.*, 1999) and could possibly explain the low esterolytic activities of FB/Bb (Kam *et al.*, 1987). A conformational

Fig. 1. Overall structure of FBSP and structure-based sequence alignment with 10 other SPs. The front view (A) and back view (B) of the overall structure in stereo ribbon diagrams are shown on the left, with corresponding surface charge potential shown on the right. The surface charge potential is plotted on a –10 to +10 scale using GRASP (Nicolls *et al.*, 1991). The labeled surface regions are colored according to distribution. The active site is designated by S195. In the structure-based sequence alignment (C) with chymotrypsin (CHM), trypsin (TRP), factor D (FD), thrombin (THM), factor Xa (FXA), factor IXa (FIXA), protein C (PRC), urokinase-type plasminogen activator (UPA), tissue-type plasminogen activator (TPA) and C2SP, the structurally equivalent positions are shaded in yellow and the secondary structure elements are underlined (β -strands, green; α -helices, blue; 3_{10} -helices, cyan). The conserved residues are colored in green; a few FBSP- or C2SP-unique residues are highlighted in red; and the potential N-linked glycosylation sites in C2SP are in blue. The variable regions are outlined in magenta and their three-dimensional positions in FBSP are shown in (A) and (B).

change to an active oxyanion hole is obviously essential for efficient catalysis.

In the absence of the positively charged N-terminus of

the SP domain, D194 forms a hydrogen bond with the hydroxyl group of S140 (Figure 2A). In C2SP, S140 is absent, but a unique R43 residue might be close enough

to form a salt bridge with E194. Other residues within the oxyanion hole, the catalytic triad and the non-specific substrate-binding site in C2SP are nearly identical to those in FBSP.

At the bottom of the S1 pocket, residue D189, typically positioning the positively charged P1 residue of the substrate for the nucleophilic attack by S195, is replaced by N189 in FBSP and S189 in C2SP. However, a typical G226, located on the opposite side to D189 at the bottom of the S1 pocket, is replaced by D226 in both enzymes (Figure 1C). Compared with the typical position of the D189 carboxyl group, the D226 carboxyl group in FBSP is more elevated (Figure 2A). Its negative charge might be more dispersed due to hydrogen bonding with the atoms from N189, T190, R225 and a non-crystallographic symmetry (NCS)-related K37. This conformation is similar to that observed in a trypsin D189G/G226D mutant, which is also less active than the wild-type trypsin (Perona *et al.*, 1993). However, relocating the negative charge from 226 to the typical 189 position in FB mutant D226N/N189D did not restore the catalytic efficiency closer to that of trypsin, as shown in our recent mutagenesis experiments (Xu *et al.*, 2000). On the contrary, this mutant exhibits complete loss of hemolytic and C3 cleavage activity and ~50% reduced activity in forming C3 convertase (Figure 3). The esterolytic activity of the mutant is also substantially reduced compared with the wild-type, mainly due to a reduction in the k_{cat} value. These data indicate that residue D226 is the primary substrate determinant for P1-Arg binding, and a proper registration of the long P1-Arg side chain to D226 on one end and its scissile bond to the nucleophilic S195 on the other end is essential for efficient catalysis (Xu *et al.*, 2000).

The S1, S2 and S3 pockets are occupied coincidentally by the side chains of K37, S36c and P36b, respectively, from an NCS-related 35 loop in a substrate-like manner (Figure 2B). The amino group of K37 forms four hydrogen bonds with oxygen atoms from residues D226, T190 and V218. The carbonyl oxygen of P36b forms a hydrogen bond with the backbone amine of G216. This crystal packing artifact results in a shift of the loop 35 tip from its typical position, but it is unlikely to affect the S1 site conformation because each structural element in the active site region is engaged in a network of specific intramolecular interactions.

The active site of FBSP displays an overall negative potential, which may provide the driving force to attract and position the natural substrates (Figure 1A). The presence of long insertion loops immediately around the active site may limit the accessibility of the FB active site to non-specific macromolecular substrates and natural SP inhibitors in blood.

Discussion

The crystal structure of FBSP reveals the presence of a unique bulky and rigid motif on the back, three distinct

deletions on the right and six loop insertions around the active site. As described, four of the six loops (loops 35, 60, 96 and 172) dictate the specificity at the S1', S2', S2 and S3 pockets. The atypical D226 residue in the S1 pocket determines P1-Arg binding and catalysis (Xu *et al.*, 2000). The remaining structural motifs, which have no corresponding regions in all other SPs with known structures, raise intriguing questions regarding their functional roles. Given that FB/Bb and C2/C2a are multidomain SPs and require cofactor binding to express proteolytic activity fully, it seems possible that these unique regions could be involved in the domain-domain interactions, cofactor binding and substrate binding.

Domain architecture

Several biophysical methods have shown that the three domains in FB constitute a compact and globular structure and that the two domains in Bb are in close contact (Smith *et al.*, 1982, 1984; Ueda *et al.*, 1987; Chamberlain *et al.*, 1998; Hinshelwood and Perkins, 1998). The linker between the vWFA domain and the SP domain is inaccessible to limited proteolysis by elastase in the context of Bb, but is susceptible in free FB (Lambris and Müller-Eberhard, 1984), suggesting a possible rearrangement of the two domains from FB to Bb.

To identify the possible domain arrangements, we performed a docking search using the FBSP crystal structure and an FB vWFA model built by Tuckwell *et al.* (1997), where the C-terminal region of the vWFA domain was restrained together with the N-terminal region of the SP domain. Although there were no clear-cut solutions, most of the best solutions placed the vWFA domain at the right of the back of the SP domain (see an example in Figure 2C). The Mg^{2+} -binding site on top of the vWFA domain and the N-terminus of the vWFA domain are exposed from the domain interface. The residues in the contact regions come from bottom loop regions of the vWFA domain and from the unique insertion motif on the back of the SP domain, suggesting that the latter probably plays a role in the interactions with the vWFA domain.

However, the exact orientation is hard to determine because the linker appears to be flexible in this structure. The N-terminal end of the SP domain could be traced to the end only for one molecule in the asymmetric unit (Table I). The C-terminal helix of the vWFA domain may also be flexible, because in isolated integrin Mac-1 I-domain (also a vWFA domain) this helix shifts in response to ligand binding (Lee *et al.*, 1995; Liddington and Bankston, 1998). Such flexibility of the linker may reflect a structural basis for the domain rearrangement mentioned above. Alternatively, it could be simply a crystal packing artifact in the isolated SP domain. On the other hand, because of the presence of potential carbohydrates at N5, N9 and N142 in C2SP, the deletion patch is unlikely to be buried in the domain interfaces. The location of the Ba

Fig. 2. Active site conformation and a model of Bb domain architecture. (A) Comparison of the essential active site structural elements in FBSP (gold) and in trypsin (silver) (catalytic triad, pink; oxyanion hole, red; non-specific substrate-binding site, green; specificity-determining residues, cyan). The unique salt bridge between D187 and K163 is in blue. A complementary ball-and-stick plot of the active site is shown in (B) together with the σ -weighted $2F_o - F_c$ electron density map (1.0 σ). Residues Lys, Ser and Pro (density in magenta) coming from an NCS-related 35 loop interact coincidentally with the S1, S2 and S3 pockets, respectively, in a substrate-like manner. (C) One of the possible positions of the vWFA domain relative to the SP domain as suggested by docking. Arrows point to the proposed cofactor-binding sites. This right-side view of the SP domain distinctly shows the distribution of the unique regions and their possible relationships with the vWFA domain and putative cofactor-binding sites.

domain in FB is uncertain, but it should be in contact with both the vWFA and SP domains (Chamberlain *et al.*, 1998; Hinshelwood and Perkins, 1998).

In other multidomain SPs, the three most common domains preceding the SP domain are the EGF domain, the Kringle domain and the CCP module (Table II).

Table II. Domain structures and cofactors of a few regulated SPs^a

SP	Cofactor protein	Domain structure
Thrombin	Thrombomodulin	Gla-2×KR-SP
Factor Xa	Factor VIIIa/Va	Gla-2×EGF-SP
Factor IXa	Factor VIIIa+Ca ²⁺	Gla-2×EGF-SP
Factor VIIa	Tissue factor	Gla-2×EGF-SP
Protein C	Protein S	Gla-2×EGF-SP
uPA	uPA receptor	EGF-KR-SP
TPA	Fibrin	FD-EGF-2×KR-SP
Plasmin	Streptokinase/staphylokinase	5×KR-SP
Tryptase	Tetramerization	SP
C1r ₂ C1s ₂	As tetramer on C1q	CUB-EGF-CUB-2×CP-SP
MASP-1,-2	MBP	CUB-EGF-CUB-2×CCP-SP
FB/Bb	C3b	3×CCP-vWFA-SP
C2/C2a	C4b	3×CCP-vWFA-SP
Factor I	Complement regulatory proteins	FIMAC-SRCR-2×LDLRA-SP

^aProteins: uPA, urokinase-type plasminogen activator; TPA, tissue-type plasminogen activator; MBP, mannose-binding protein; MASP, MBP-associated serine protease. Domains or modules: Gla, γ -carboxy glutamic acid-rich domain; KR, Kringle domain; EGF, epidermal growth factor domain; FD, finger domain; CUB, C1r/C1s, sea urchin protein Uegf, and bone morphogenetic protein-1; CCP, complement control protein; FIMAC, factor I-membrane attack complex proteins C6/C7 module; SRCR, scavenger receptor Cys-rich module; LDLRA, low-density lipoprotein receptor class A.

However, only a few multidomain SP structures containing the EGF domains currently are available, e.g. factor Xa, factor IXa, factor VIIa and protein C (reviewed in Bode *et al.*, 1997). Superposition of these structures onto FBSP showed that a similar C1–C122 disulfide bond is utilized to link the SP domain with its preceding domain. Similarly, the regions on the back of these SPs also mediate the domain interface. However, the FBSP- and C2SP-unique helix (E125b–A125g) in the motif on the back overlaps significantly with the core of the proximal EGF domain, suggesting that the vWFA–SP interface would be different from the EGF–SP interface.

This attachment on the back of other types of domains could be a general feature in the multidomain SPs. Such a domain arrangement would allow the other domains to facilitate the regulation of the SP domain through binding various cofactors without hindering the accessibility of the active site in the front.

Cofactor binding and possible activation mechanisms

Various biochemical experiments have demonstrated the presence of cofactor-binding sites in each of the three domains (reviewed in Volanakis, 1988). The binding site on the vWFA domain has been located by mutagenesis experiments to be around the Mg²⁺-binding motif of FB and C2 (Horiuchi *et al.*, 1991; Hourcade *et al.*, 1995; Tuckwell *et al.*, 1997). The binding site on the SP domain of FB was also suggested for a 33 kDa fragment corresponding to FBSP based on its C3b-binding activity (Lambris and Müller-Eberhard, 1984) and limited C3 cleavage activity that is independent of Mg²⁺ but dependent on C3b (Sanchez-Corral *et al.*, 1990). This fragment behaves like Bb in the absence of Mg²⁺ in that they have similar proteolytic activity and their cofactor-assisted proteolysis can be inhibited similarly by factor H, whereas,

Fig. 3. Relative cobra venom factor (CoVF)-binding activity of FB mutants compared with the wild-type FB or Bb. The methods are described in Xu *et al.* (1999).

in the presence of Mg²⁺, Bb has enhanced proteolytic activity, suggesting that the vWFA domain assists C3b binding (Sanchez-Corral *et al.*, 1990). Additionally, some mutations at residues D226, D187 and N189 in the SP domain were shown to reduce the binding of FB and Bb to a C3b analog, cobra venom factor (Xu *et al.*, 2000; Figure 3). However, the presence of the cofactor-binding sites on both domains is not supported by an early electron microscopy visualization of the bimolecular complexes, suggesting that in most populations only one blob attached the cofactor protein (Smith *et al.*, 1984). Given that the half-life of the bimolecular complexes is <2 min in solution (Kerr, 1980; Pangburn and Müller-Eberhard, 1986), the interaction between the SP domain and the cofactor protein can not be ruled out.

From the structure of FBSP, the cofactor-binding site on the SP domain can be postulated based on the unique regions where FBSP and C2SP differ, because FB/Bb binds C3b but C2/C2a binds C4b. Two regions stand out from such comparisons. One region is the deletion patch, which in FBSP displays mostly charged residues, but in C2 displays neutral and hydrophilic residues. This patch is bordered by the C-terminal end of the linker region on the right and loops 35, 186 and 218 on the left (Figure 2C), all of which are unique to but different between FBSP and C2SP (Figure 1C). The deletions or shortenings of loops 16, 71 and 142, which typically mediate zymogen activation, appear to leave a space for something to access and activate the oxyanion hole. This flat and hydrophilic surface also fits the predominant features suitable for weak hetero protein–protein interactions (Jones and Thornton, 1996) and seems capable of accommodating a large molecule such as C3b or C4b (~200 kDa). However, the potential presence of a glycan on the 142 loop of C2SP would alter its surface features, which may or may not exclude C4b binding.

Another region where FBSP and C2SP differ is underneath or in front of the S1 pocket, near the 164 insertion, 186 loop and 218 loop (Figure 2C). This region features

a constrained *cis* conformation for P186b in FBSP (absent in C2SP), a unique buried salt bridge between K163 and D187, flexible conformations for the middle part of the 218 loop (seven residues longer in C2SP) and a few exposed hydrophobic residues. Mutations such as D187Y, N189A/D187A, D226G and D226N/N189D caused ~50% reduction of the cofactor-binding activity of Bb, and mutation of D226G/N189D caused 87% reduction (Figure 3) (Xu *et al.*, 2000). These mutations also reduced the cofactor-binding activity of FB proportionally, indicating that these residues or their exposed neighboring residues might contribute to C3b binding for both FB and Bb. However, without knowledge of the local conformational changes due to the above substitutions in the context of FB or Bb, we can not be certain of their roles in C3b binding.

Both sites are close enough to the active site to allow direct regulation of the active site conformation by the cofactor (Figure 2C). It is also possible that they act together in binding the large cofactor molecule. The Mg^{2+} -binding site on the vWFA, although not close to these two sites, can still contribute to cofactor binding if the large size of the cofactor is considered. In view of the physical separation between the atypical active site in the SP domain and the Mg^{2+} -binding site in the vWFA domain (Figure 2C), it is difficult to visualize the activation of the former as being entirely due to the translation of a cross-domain conformational change. The observed clustering of the atypical oxyanion hole, deletion patch and 218 loop, probably present in Bb, might suggest a possible role and site for cofactor binding in the SP domain.

The cofactor may either activate the oxyanion hole or act as a platform for binding the natural substrate. The latter is supported by the fact that although cofactor binding is essential for proteolytic activity, it did not enhance the esterolytic activity toward a C3-like tripeptide substrate (Ikari *et al.*, 1983). If so, then the oxyanion hole could be activated by the natural substrate, C3. C3b, generated by either C3b-bound Bb or C4b-bound C2a, is another bulky molecule that may need to be accommodated in the proximity of the FBSP active site for efficient proteolysis of C5. It would be reasonable to assume that the deletion patch acts as a substrate-binding platform, while the other site below or in front of the S1 pocket assists in the binding of cofactor and the substrates C3 and C5. Obviously more functional and structural characterizations are necessary in order to differentiate these possibilities and to pinpoint the cofactor-binding sites.

In other SPs where cofactor binding is required for activation, the cofactor could approach the SP domain from either the right side (loops 16, 71 and 142), as in the case of the plasmin–streptokinase complex (Wang *et al.*, 1998), or from the bottom (residues 125–134 and 164–170b) as in the case of the factor VIIa–tissue factor complex (Banner *et al.*, 1996). However, the N-terminus of the SP domain in both cases is free and tucked inside, thus also playing a substantial role in activation, which is not possible in Bb/C2a.

Taken together, the activation mechanism of complement FB and C2 further broadens our current understanding of SP regulation mechanisms. The structural element responsible for SP regulation can be as small and

efficient as a single residue insertion, such as Arg218 in complement factor D (Volanakis and Narayana, 1996; Jing *et al.*, 1998), or as large and elaborate as the requirements of other domains and cofactors, such as the EGF domains and tissue factor for factor VIIa (Banner *et al.*, 1996) or the vWFA domain and C3b/C4b cofactor for Bb/C2a. More diversity may be expected for other regulated SPs, depending on the specific control needed.

Materials and methods

Protein expression, purification and crystallization

The cDNA sequence coding for FBSP was amplified by PCR using the FB cDNA clone BHL4-1 as template. The PCR product was restricted by *Bam*HI and *Eco*RI and subcloned into a baculovirus transfer vector, pAcGP67-A (PharMingen, San Diego, CA). Transfection and expression of FBSP in baculovirus-infected Sf9 cells were carried out according to the manufacturer's protocols. The recombinant FBSP consists of a vector-derived tripeptide Ala–Asp–Pro at the N-terminus and 295 C-terminal amino acid residues of FB. The protein was purified by ion exchange chromatography using a CM–Sepharose column followed by a Mono S column (Amersham Pharmacia Biotech, Piscataway, NJ). The starting buffer for the CM–Sepharose column is 0.1 M sodium acetate, 0.02 M EDTA (pH 6.3), and for the Mono S column is 0.05 M phosphate (pH 6.8). The elution buffer contains a linear gradient of the starting buffer and 1.0 M NaCl. Purified protein was dialyzed against 20 mM Tris, 100 mM NaCl (pH 7.5) and concentrated to 20 mg/ml. Crystals were obtained at room temperature by the hanging drop vapor diffusion method. The reservoir solution contained 35% PEG-1000 and 30 mM MES (pH 6.0), and the drops contained an equal volume of reservoir solution and protein solution.

Data collection and phasing

Two native data sets were collected on in-house X-ray generators (40 kV, 100 mA), one at room temperature on a RAXIS-II detector, and the other one at 95 K on a DIP detector using the mother liquor as cryoprotectant, both processed using the DENZO and SCALEPACK programs (Otwinowski and Minor, 1997). The two data sets were non-isomorphous as the *b*-axis differed by 1.7 Å (Table I). The crystals belonged to space group *P*2₁ with two molecules in the asymmetric unit. While searching for molecular replacement solutions, we also collected heavy atom derivative data at both room and low temperatures (Table I). The positions of the heavy atoms were identified and refined using the XTALVIEW (McRee, 1992) and MLPHARE (in CCP4, 1994) programs. Later, molecular replacement solutions were found in the AMoRe program (Navaza, 1994) using an SP model containing only the core β -sheets. The phases calculated using these solutions confirmed the heavy atom sites on the difference Fourier maps and brought the sites from different derivatives to one relative origin. Next, 4-fold multiple crystal averaging was performed using the DMMULTI program (CCP4, 1994), which not only combined the phases from independent sources but also extended the phases to the highest resolution in the native data (reviewed in Kleywegt and Read, 1997). This averaged map helped to interpret the positions of ~90% of the residues.

Structure refinement

The structure was refined using the CNS program with its maximum likelihood-targeted torsion angle dynamics, positional refinement and *B*-factor refinement protocols (Brünger *et al.*, 1998). Except for 10% reflections excluded for *R*_{free} calculation (Brünger, 1992; Kleywegt and Brünger, 1996), all reflections from 20.0 to 2.1 Å were used in the refinement. Sigma-weighted electron density maps (Read, 1986) were calculated using CNS and guided manual rebuilding on O (Jones *et al.*, 1991). Structure quality (Table I) was checked in each cycle using OOPS (Kleywegt and Jones, 1996, 1997), and later with CNS, PROCHECK (Laskowski *et al.*, 1993) and WHAT_CHECK (Hooft *et al.*, 1996).

Structural analyses and docking

Structure-based alignments were obtained using MODELLER (Sali and Blundell, 1993) and O. The C2SP sequence was aligned with the FBSP sequence and a model was built with MODELLER. The docking search was performed using SoftDock implemented by Mike Carson on the basis of methods described in Jiang and Kim (1991). Unless specified, structural figures were prepared using RIBBONS (Carson, 1997). The

structure and diffraction data are deposited in the PDB with the accession code 1DLE.

Acknowledgements

The authors are grateful to the staff at the Center for Macromolecular Crystallography (CMC) for their enormous support in using the state of the art X-ray crystallography and computational facilities. H.J. deeply appreciates Yu Luo, Todd Bice and Champion Deivanayagam (CMC) for discussions on the MIR method; Kevin Cowtan (York University, UK) for suggestions on multiple crystal averaging; Danny Tuckwell (University of Manchester, UK) for the model of the FB vWFA domain; Xuejun Zhang (Oklahoma Medical Research Foundation) for C_{α} coordinates of the plasmin–streptokinase complex before release; and Lei Jin and Timothy Springer (Harvard University) for generous support, help in computer usage and enlightening discussions. The project is funded by NIH grants AI 39818 to S.V.L.N. and AR 44506 to J.E.V.

References

- Arlaud, G.J., Volanakis, J.E., Thielens, N.M., Narayana, S.V.L., Rossi, V. and Xu, Y. (1998) The atypical serine proteases of the complement system. *Adv. Immunol.*, **69**, 249–307.
- Banner, D.W., D'Arcy, A., Chene, C., Winkler, F.K., Guba, A., Konigsberg, W.H., Nemerson, Y. and Kirchhofer, D. (1996) The crystal structure of complex of blood coagulation factor VIIa with soluble tissue factor. *Nature*, **380**, 41–46.
- Bode, W. and Huber, R. (1992) Natural protein proteinase inhibitors and their interaction with proteinases. *Eur. J. Biochem.*, **204**, 433–451.
- Bode, W., Brandstetter, H., Mather, T. and Stubbs, M.T. (1997) Comparative analysis of haemostatic proteinases: structural aspects of thrombin, factor Xa, factor IXa and protein C. *Thromb. Haemostasis*, **78**, 501–511.
- Brünger, A.T. (1992) Free R value: a novel statistical quantity for assessing the accuracy of crystal structure. *Nature*, **355**, 472–475.
- Brünger, A.T. et al. (1998) Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D*, **54**, 905–921.
- Carson, M. (1997) Ribbons. *Methods Enzymol.*, **277**, 493–505.
- CCP4 (Collaborative Computational Project Number 4, 1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D*, **50**, 760–763.
- Chamberlain, D., Hinshelwood, J. and Perkins, S.J. (1998) The compact five-domain structure of factor B of human complement by X-ray and neutron solution scattering and homology modeling. *Mol. Immunol.*, **35**, 390.
- Czapinska, H. and Otlewski, J. (1999) Structure and energetic determinants of the S1-site specificity in serine proteases. *Eur. J. Biochem.*, **260**, 571–595.
- Fishelson, Z. and Müller-Eberhard, H.J. (1984) Residual hemolytic and proteolytic activity expressed by Bb after decay-dissociation of C3b,Bb. *J. Immunol.*, **132**, 1425–1429.
- Hinshelwood, J. and Perkins, S.J. (1998) The solution structure of human factor B reveals interactions between its domains: a study by NMR and homology modeling. *Mol. Immunol.*, **35**, 389.
- Hoof, R.W., Vriend, G., Sander, C. and Abola, E.E. (1996) Errors in protein structures. *Nature*, **381**, 272–275.
- Horiuchi, T., Macon, K.J., Engler, J.A. and Volanakis, J.E. (1991) Site-directed mutagenesis of the region around Cys-241 of complement component C2. *J. Immunol.*, **147**, 584–589.
- Hourcade, D.E., Wagner, L.M. and Oglesby, T.J. (1995) Analysis of the short consensus repeats of human complement factor B by site-directed mutagenesis. *J. Biol. Chem.*, **270**, 19716–19722.
- Huber, R. and Bode, W. (1978) Structural basis of the activation and action of trypsin. *J. Am. Chem. Soc.*, **100**, 114–122.
- Ikari, N., Hitomi, Y., Ninobe, M. and Fujii, S. (1983) Studies on esterolytic activity of alternative complement component factor B. *Biochim. Biophys. Acta*, **742**, 318–323.
- Jiang, F. and Kim, S.-H. (1991) 'Soft docking': matching of molecular surface cubes. *J. Mol. Biol.*, **219**, 79–102.
- Jing, H., Babu, Y.S., Moore, D., Kilpatrick, J.M., Liu, X.-Y., Volanakis, J.E. and Narayana, S.V.L. (1998) Structures of native and complexed complement factor D: implications of the atypical His57 conformation and self-inhibitory loop in the regulation of specific serine protease activity. *J. Mol. Biol.*, **282**, 1061–1081.
- Jing, H., Macon, K.J., Moore, D., DeLucas, L.J., Volanakis, J.E. and Narayana, S.V.L. (1999) Structural basis of profactor D activation: from a highly flexible zymogen to a novel self-inhibited serine protease, complement factor D. *EMBO J.*, **18**, 804–814.
- Jones, S. and Thornton, J.M. (1996) Principles of protein–protein interactions. *Proc. Natl Acad. Sci. USA*, **93**, 13–20.
- Jones, T.A., Zou, J.Y., Cowan, S.W. and Kjeldgaard, M. (1991) Improved methods for binding protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A*, **47**, 110–119.
- Kam, C.-M., McRae, B.J., Harper, J.W., Niemann, M.A., Volanakis, J.E. and Powers, J.C. (1987) Human complement proteins D, C2 and B. Active site mapping with peptide thioester substrates. *J. Biol. Chem.*, **262**, 3444–3451.
- Kerr, M.A. (1980) The human complement system. Assembly of the classical pathway C3 convertase. *Biochem. J.*, **189**, 173–181.
- Khan, A.R. and James, M.N.G. (1998) Molecular mechanisms for the conversion of zymogens to active proteolytic enzymes. *Protein Sci.*, **7**, 815–836.
- Kleywegt, G.J. and Brünger, A.T. (1996) Checking your imagination: applications of the free R value. *Structure*, **4**, 897–904.
- Kleywegt, G.J. and Jones, T.A. (1996) Efficient rebuilding of protein structures. *Acta Crystallogr. D*, **52**, 829–832.
- Kleywegt, G.J. and Jones, T.A. (1997) Model rebuilding and refinement practice. *Methods Enzymol.*, **276**, 208–230.
- Kleywegt, G.J. and Read, R.J. (1997) Not your average density. *Structure*, **5**, 1557–1569.
- Kraut, J. (1977) Serine proteases: structure and mechanism of catalysis. *Annu. Rev. Biochem.*, **46**, 331–358.
- Lambris, J.D. and Müller-Eberhard, H.J. (1984) Isolation and characterization of a 33,000-dalton fragment of complement factor B with catalytic and C3b binding activity. *J. Biol. Chem.*, **259**, 12685–12690.
- Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structure. *J. Appl. Crystallogr.*, **26**, 283–291.
- Lee, J.-O., Bankston, L.A., Arnaout, M.A. and Liddington, R.C. (1995) Two conformations of the integrin A-domain (I-domain): a pathway for activation? *Structure*, **3**, 1333–1340.
- Liddington, R. and Bankston, L. (1998) The integrin I domain: crystals, metals and related artefacts. *Structure*, **6**, 937–938.
- McRee, D.E. (1992) A visual protein crystallographic software system for X11/XView. *J. Mol. Graphics*, **10**, 44–46.
- Navaza, J. (1994) AMoRe: an automated package for molecular replacement. *Acta Crystallogr. A*, **50**, 157–163.
- Nicolls, A., Sharp, K.A. and Honig, B. (1991) Protein folding and association: insights from the interfacial and thermodynamics properties of hydrocarbons. *Proteins Struct. Funct. Genet.*, **11**, 281–296.
- Otwiński, Z. and Minor, W. (1997) Processing X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307–326.
- Pangburn, M.K. and Müller-Eberhard, H.J. (1986) The C3 convertase of the alternative pathway of human complement. *Biochem. J.*, **235**, 723–730.
- Pereira, P.J.B., Berger, A., Macedo-Ribeiro, S., Huber, R., Matschner, G., Fritz, H., Sommerhoff, C.P. and Bode, W. (1998) Human β -trypsin is a ring-like tetramer with active sites facing a central pore. *Nature*, **392**, 306–311.
- Perona, J.J. and Craik, C.S. (1995) Structural basis of substrate specificity in the serine proteases. *Protein Sci.*, **4**, 337–360.
- Perona, J.J., Tsu, C.A., McGrath, M.E., Craik, C.S. and Fletterick, R.J. (1993) Relocating a negative charge in the binding pocket of trypsin. *J. Mol. Biol.*, **230**, 934–949.
- Read, R.J. (1986) Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallogr. A*, **42**, 140–149.
- Sali, A. and Blundell, T.L. (1993) Comparative protein modeling by satisfaction of spatial restraints. *J. Mol. Biol.*, **234**, 779–815.
- Sanchez-Corral, P., Anton, L.C., Alcolea, J.M., Marques, G., Sanchez, A. and Vivanco, F. (1990) Proteolytic activity of the different fragments of factor B on the third components of complement (C3). Involvement of the N-terminal domain of Bb in magnesium binding. *Mol. Immunol.*, **27**, 891–900.
- Smith, C.A., Vogel, C.-W. and Müller-Eberhard, H.J. (1982) Ultrastructure of cobra venom factor-dependent C3/C5 convertase and its zymogen, factor B of human complement. *J. Biol. Chem.*, **257**, 9879–9882.
- Smith, C.A., Vogel, C.-W. and Müller-Eberhard, H.J. (1984) An electron microscopy study of the C3 convertase of human complement. *J. Exp. Med.*, **159**, 324–329.
- Tuckwell, D.S., Xu, Y., Newham, P., Humphries, M.J. and Volanakis, J.E. (1997) Surface loops adjacent to the cation-binding site of the

- complement factor B–von Willebrand factor type A module determine C3b binding specificity. *Biochemistry*, **36**, 6605–6613.
- Ueda,A., Kearney,J.F., Roux,K.H. and Volanakis,J.E. (1987) Probing functional sites on complement protein B with monoclonal antibodies: evidence for C3b-binding sites on Ba. *J. Immunol.*, **138**, 1143–1149.
- Volanakis,J.E. (1988) C3 convertase of complement: molecular genetics, structure and function of the catalytic domains, C2 and B. *Year Immunol.*, **4**, 218–230.
- Volanakis,J.E. and Narayana,S.V.L. (1996) Complement factor D, a novel serine protease. *Protein Sci.*, **5**, 553–564.
- Wang,X., Lin,X., Loy,J.A., Tang,J. and Zhang,X.C. (1998) Crystal structure of the catalytic domain of human plasmin complexed with streptokinase. *Science*, **281**, 1662–1665.
- Westel,R.A., Kulies,J., Lökk,M.-L., Kiepiela,P., Akama,H., Johnson,C.A.C., Densen,P. and Colten,H.R. (1996) Type II human complement C2 deficiency. Allele-specific amino acid substitutions (Ser189 to Phe, Gly444 to Arg) cause impaired C2 secretion. *J. Biol. Chem.*, **271**, 5824–5831.
- Xu,Y. and Volanakis,J.E. (1999) C2. In Mooley,J.B. and Walport,M.J. (eds), *FactsBook of Complement System*. Academic Press, San Diego, CA, pp. 73–77.
- Xu,Y., Circolo,A., Jing,H., Wang,Y., Narayana,S.V.L. and Volanakis,J.E. (2000) Mutational analysis of the primary substrate specificity pocket of complement factor B: Asp226 is a major structural determinant for P1-Arg binding. *J. Biol. Chem.*, in press.

Received August 9, 1999; revised and accepted November 3, 1999