REGULAR ARTICLE

Structures of human plasma β–factor XIIa cocrystallized with potent inhibitors

Alexey Dementiev,[1] Abel Silva,[2] Calvin Yee,[2] Zhe Li,[2] Michael T. Flavin,[1] Hing Sham,[2] and James R. Partridge[2]

[1]Shamrock Structures, LLC, Woodridge, IL; and [2]Global Blood Therapeutics, Inc, South San Francisco, CA

<table>
<tr><td>

Key Points

- The first crystal structure of human plasma β-FXIIa in its active state is presented. The conformational lability of FXIIa is discussed.

- These novel structural data provide molecular insight into β-FXIIa interaction with its substrates and inhibitors.

</td></tr>
</table>

Activated factor XIIa (FXIIa) is a serine protease that has received a great deal of interest in recent years as a potential target for the development of new antithrombotics. Despite the strong interest in obtaining structural information, only the structure of the FXIIa catalytic domain in its zymogen conformation is available. In this work, reproducible experimental conditions found for the crystallization of human plasma β-FXIIa and crystal growth optimization have led to determination of the first structure of the active form of the enzyme. Two crystal structures of human plasma β-FXIIa complexed with small molecule inhibitors are presented herein. The first is the noncovalent inhibitor benzamidine. The second is an aminoisoquinoline containing a boronic acid–reactive group that targets the catalytic serine. Both benzamidine and the aminoisoquinoline bind in a canonical fashion typical of synthetic serine protease inhibitors, and the protease domain adopts a typical chymotrypsin-like serine protease active conformation. This novel structural data explains the basis of the FXII activation, provides insights into the enzymatic properties of β-FXIIa, and is a great aid toward the further design of protease inhibitors for human FXIIa.

Introduction

Human factor XIIa (FXIIa; Hageman factor, EC 3.4.21.38), a multidomain serine protease of the trypsin-like family, initiates the intrinsic coagulation cascade by contact activation in a reaction involving high-molecular-weight kininogen (HMWK) and plasma prekallikrein (PPK).[1] This activation requires proteolytic conversion of plasma FXII zymogen to active protease FXIIa on negatively charged surfaces where FXII undergoes conformational changes and small amounts of active FXIIa are formed.[2] At the same time, HMWK bound to the same surface presents PPK to FXIIa for activation. The resulting active plasma kallikrein (PK) reciprocally activates additional FXIIa in a positive feedback loop.[3] In the next steps of the intrinsic pathway, FXIIa cleaves its substrate FXI to generate active FXIa, which in turn activates FIX to FIXa.[4,5] This series of reactions eventually drives thrombin generation and fibrin formation in the final steps of coagulation. FXII deficiency in humans and animals is not associated with excessive bleeding, demonstrating that FXIIa activation of FXI is not essential for hemostasis.[6,7] Except for procoagulant activity, the FXIIa-driven contact system has proinflammatory activity via the kallikrein-kinin system, which liberates the inflammatory mediator bradykinin from HMWK via PK.[8,9] FXIIa activity in plasma is mainly regulated by its cognate serpin C1 esterase inhibitor (C1INH).[10] Plasma antithrombin (AT) and plasminogen activator inhibitor 1 (PAI-1) also have some minor FXIIa inhibitory activity.[11]

Thus, recent data have made FXIIa an attractive target for designing safe anticoagulants that inhibit thrombosis without the influence of hemostasis. Currently available antithrombotic agents such as low-molecular-weight heparins, warfarin, and antiplatelet therapies are associated with a high risk of severe bleeding complications because they target components of the blood-clotting mechanism such as

thrombin, FVIIa, FIXa, FXa, and FXIa.[12] Therefore, designing new drugs against FXIIa, which is involved in the development of pathological thrombus formation while having limited effect on physiological homeostasis, may make antithrombotic therapy safer. However, this strategy is currently limited by the absence of structural data for active FXIIa. Recently, the crystal structure of the catalytic domain of recombinant human FXII (residues 354-596, FXII mature protein sequencing, FXIIc) was determined.[13] The structure revealed the zymogen conformation of the enzyme catalytic domain and did not provide a suitable platform for a structure-based drug design approach.

The inactive zymogen form of FXII is secreted as a single-chain polypeptide chain of 596 amino acid residues with a molecular weight of 80 kDa. Upon contact system activation, surface-bound FXII proenzyme is cleaved at the Arg353-Val354 peptide bond by PK, generating α-FXIIa consisting of a 50-kDa heavy chain and a 28-kDa light chain held together by the Cys340-Cys467 disulfide bond. Further proteolytic cleavages of α-FXIIa heavy chain at the Arg334 and Arg343 C termini by PK yields β-FXIIa, consisting of 2 polypeptide chains of a 2-kDa heavy chain remnant and a 28-kDa catalytic domain covalently bonded together by the same disulfide bond.[14]

In this work, we crystallized and solved, for the first time, structures of human plasma β-FXIIa in complex with 2 different inhibitors. These include the noncovalent inhibitor benzamidine, a classical inhibitor for serine proteases, and a covalent synthetic small molecule inhibitor containing a surrogate of the basic group. For convenience in structural comparison with other members of the trypsin-like serine protease family, we used the chymotrypsinogen residue numbering shown in italics (supplemental Figure 1).[14] The catalytic domain of β-FXIIa adopts the same fold typical of other active serine proteases. Structural analysis of surface electrostatic potentials revealed a plausible exosite, which may be crucial for the interactions of cofactors, substrates, and inhibitors with β-FXIIa. Also, the structures of FXIIa-inhibitor complexes provide important information about the geometry of the binding site, which is a critical asset for understanding the determinants for selectivity and specificity in FXIIa-ligand interactions.

Methods

Materials

Human plasma β-FXIIa and PK were purchased from Molecular Innovations (Novi, MI). The purity of the commercial β-FXIIa sample was 95% as assessed by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) analysis (Molecular Innovations) and electrospray mass spectroscopy (supplemental Figure 2). FXIa and Lys-plasmin were obtained from Hematologic Technology (Burlington, VT), and bovine trypsin was obtained from Worthington Biochemical Corporation (Lakewood, NJ). The chromogenic substrates D-Pro-Phe-ArgpNA (S-2302) and N-Z-D-Arg-Gly-Arg-pNA (S-2765) were obtained from Aniara (Westchester, OH); tosyl-Gly-Pro-Lys-4-nitranilide (Chromozyme PL) was obtained from Sigma-Aldrich (St. Louis, MO). Synthesis of compound 1 (3-(1-aminoisoquinolin-6-yl)phenyl)boronic acid), the protease inhibition assay, and the modeling of β-FXIIa interactions with its inhibitor and substrate are described in detail in supplemental Methods.

Crystallization and structure determination of human plasma β-FXIIa

Human plasma β-FXIIa protein with the presence of 8 mM benzamidine was concentrated to 10 mg/mL using Amicon centrifugal filters with a 10-kDa molecular-weight cutoff (Millipore) in 30 mM N-2-hydroxyethylpiperazine-N′-2-ethanesulfonic acid (HEPES) buffer, pH 7.0. Initial crystals of the β-FXIIa–benzamidine complex were grown at 18°C from drops containing 0.3 μL of the protein sample and 0.3 μL of reservoir solution (0.2 M ammonium iodide, 2.2 M ammonium sulfate). Protein crystals of the complex suitable for x-ray analysis were obtained after several cycles of microseeding under similar crystallization conditions. SDS-PAGE analysis (supplemental Figure 2B) of protein samples obtained by dissolving the crystals in SDS buffer did not reveal any degradation products and confirmed the presence of only β-FXIIa protein in the crystals used for the data collection.

A soaking approach was used to obtain the protein crystals of the compound 1–β-FXIIa complex. The benzamidine crystals were transferred to a drop of mother liquor containing a saturated solution of compound 1 ligand. The drop was further incubated for 18 hours at 18°C over the reservoir solution before the data collection.

For data collection, crystals were harvested with 20% (vol/vol) glycerol in the reservoir solution. Diffraction data were collected from a single flash-frozen crystal on the LS-CAT ID beamline (APS; Argonne National Laboratory). Data were indexed and processed with HKL-2000.[15] The crystals belonged to tetragonal space group I41 and contained 1 molecule of β-FXIIa per asymmetric unit.

The structure of β-FXIIa was solved by molecular replacement using the PHASER program from the CCP4 software suite, with the structure of bovine trypsin (Protein Data Bank [PDB] code 3MFJ) as a search model.[16,17] The final structure of β-FXIIa was obtained by carrying out several cycles consisting of manual model building using COOT, followed by structure refinement with REFMAC from the CCP4 software suite.[18,19] Coordinates have been deposited in the PDB (codes 6B74 and 6B77).

Results

Overall structure of human plasma β-FXIIa

The structures of human plasma β-FXIIa in complex with the small molecule compound 1 (3-(1-aminoisoquinolin-6-yl)phenyl)boronic acid) and benzamidine were solved at a resolution of 2.4 Å and 2.3 Å, respectively (Table 1). For both structures, all 9 residues (residues 335-343, Asn5-Arg4) of the heavy chain remnant and 252 residues (residues 354-596, Val16-Ser244) of the FXIIa light chain, with the exception of a few side chains, were traceable in the final electron-density map (Figure 1). These 2 crystal structures are very similar to one another with an overall root mean square deviation (RMSD) of 0.25 Å for all Cα pairs (Figure 1C). The protease domain of β-FXIIa displays strong structural similarity with other trypsin-like active serine proteases, consisting of 2 β-barrels forming the substrate-binding site and the catalytic triad located at the cleft between 2 barrels in a typical arrangement (Figure 1A). The short remnant heavy chain is positioned on the molecular surface opposite to the active site entrance, and is covalently linked to the protease domain through disulfide bridge Cys1-Cys122. The remnant chain also makes contacts from both the backbone and

Table 1. Data collection and refinement statistics

PDB code	Benzamidine complex 6B74	Compound 1 complex 6B77
Data collection		
Wavelength, Å	0.9795	0.9787
Resolution range, Å	50.00-2.32 (2.40-2.32)	50.00-2.37 (2.47-2.37)
Space group	I 41	I 41
Unit cell: a, b, c, Å	80.289, 80.289, 121.731	79.402, 79.402, 122.246
Total reflections	102 197	95 803
Unique reflections	16 576 (1 632)	15 282 (1 524)
Multiplicity	6.2 (6.4)	6.3 (6.4)
Completeness, %	99.9 (100.0)	99.8 (100.0)
$<I>/\sigma(I)$	18.19 (2.98)	21.4 (3.4)
$Rmerge$ (%)*	8.4 (84.9)	7.9 (86.7)
Refinement		
Resolution range, Å	36.22-2.32 (2.41-2.32)	41.35-2.37 (2.45-2.37)
$Rwork$ (%)/$Rfree$ (%)†	19.0 (29.6)/22.2 (32.7)	17.4 (24.4)/21.8 (32.2)
No. of nonhydrogen atoms‡	1952	2015
Protein	1849	1848
Ligands	64	70
Waters	51	97
RMS bonds, Å	0.004	0.012
RMS angles, °	0.832	1.360
Ramachandran favored, %	96.7	95.6
Ramachandran allowed, %	2.5	3.6
Ramachandran outliers, %	0.8	0.8
Average B factor, Å²	71.8	58.0
Macromolecules	71.2	57.4
Ligands	92.5	73.2
Waters	67.4	58.1

Statistics for the highest-resolution shell are shown in parentheses.

RMS, root mean square deviation from ideal values (crystallography).

*$Rmerge = 100\Sigma(h)\Sigma(i)|I(i)-<I>|/\Sigma(h)\Sigma(i)I(i)$, where $I(i)$ is the ith intensity measurement of reflection h, and $<I>$ is the average intensity from multiple observations.

†$Rfactor = \Sigma||Fobs|-|Fcalc||/\Sigma|Fobs|$. Where $Fobs$ and $Fcalc$ are the structure factor amplitudes from the data and the model, respectively; 10% reflections were used to calculate $Rfree$ values.

‡Per asymmetric unit.

side chains to residues in the protease domain. These include a hydrogen bond from the backbone carbonyl of *Ser-1* to the backbone amino group of *Arg205C*, a hydrogen bond from the carbonyl of *Gly2* to the backbone amino group of *Glu205B*, and a hydrogen bond from the backbone amino group of *Arg4* to the carbonyl oxygen atom of *Ala205*. Finally, the side chain of *Arg4* forms a hydrogen bond with the backbone carbonyl of *Gln204* and the NH1 group of *Arg205C* with the carbonyl of *Leu-2*. It is interesting to note that the remnant chain of β-FXIIa is organized in a conformation different to that observed for human tissue-type plasminogen activator (tPA), human urokinase-type plasminogen activator (uPA), thrombin, or human hepatocyte growth factor activator (HGFA), although this chain is also linked to the protease domain by a homologous disulfide bridge, but has a much shorter length at its C terminus (supplemental Figure 1). Optimal superposition of the FXIIa catalytic domain with corresponding HGFA, tPA, and uPA homologous domains results in 237, 232, and 231

equivalent Cα atoms with the RMSD values of 1.26 Å, 1.45 Å, and 1.50 Å, respectively.

The protease domain of human FXIIa has 6 disulfide bonds (Figure 1B). Three of them, *Cys42-Cys58*, *Cys168-Cys182*, and *Cys191-Cys220* are well conserved among trypsin-like proteases (supplemental Figure 1). Two other disulfide bonds, *Cys50-Cys111* and *Cys136-Cys201*, are less observed. The first of these 2 bridges is found in the structures of HGFA, tPA, and uPA. The second bridge, homologous to *Cys136-Cys201*, occurs in the protease domains of HGFA, tPA, and uPA, and also in trypsin and chymotrypsinogen. Finally, there is 1 disulfide bond, *Cys77-Cys80*, which is characteristic for human plasma FXIIa.

There is 1 potential N-glycosylation site in the molecule of β-FXIIa and mass spectroscopic analysis of the commercial sample confirmed the presence of 2911-Da oligosaccharides covalently linked to the protein moiety (supplemental Figure 2A). This
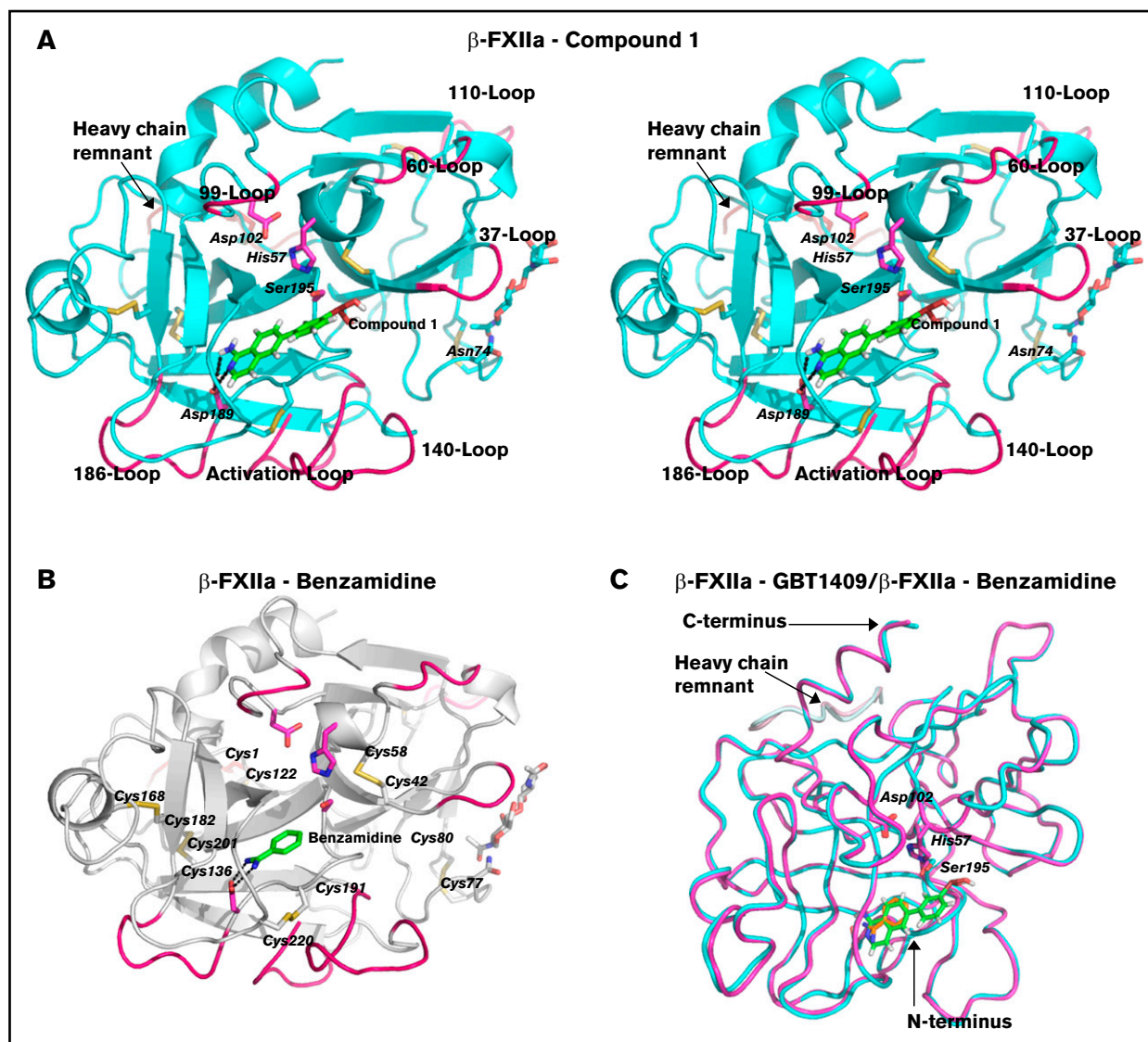
Figure 1. Crystal structure of human plasma β-FXIIa. (A) Stereo ribbon diagram (cyan) of β-FXIIa in complex with covalent compound 1 inhibitor in "standard" orientation. Compound 1 is shown in green sticks. The side chains of key β-FXIIa residues are shown as magenta sticks: *His57*, *Asp102*, and *Ser195* of the catalytic triads, *Asp189* at the base of the S1 specificity pocket. The side chains of N-glycosylated *Asn74* and NacGlc-disaccharide are shown in cyan sticks. Some surface loops discussed in the text are labeled and shown in pink. The heavy chain remnant of β-FXIIa is shown in red. Hydrogen bonds are indicated by black dotted lines. (B) The structure of β-FXIIa in complex with noncovalent benzamidine inhibitor is represented as in panel A. Six disulfide bonds are indicated and are shown in yellow. (C) Superposition of the β-FXIIa (benzamidine [magenta] and β-FXIIa) compound 1 (cyan) crystal structures. The catalytic domains are superimposed by aligning the Cα atoms of residues *16-244* and are presented as cartoon diagrams. Benzamidine (orange), compound 1 (green), and residues of the catalytic triad are shown in a stick representation.

N-glycosylation site at *Asn74* is located on the surface loop, which is spatially removed from the active site cleft and the substrate-binding site (Figure 1A-B), and, most likely, the glycosylation should not affect FXIIa catalytic properties, as observed for other glycosylated active serine proteases such as PK and FXIa.[20,21]

Human plasma β-FXIIa contains 26 negatively charged residues and 15 positively charged residues (and 9 histidine residues), and is a relatively acidic protein, with a calculated isoelectric point value of 5.2, among trypsin-like serine proteases. Analysis of surface charge calculated from the crystal structure of the covalent complex between β-FXIIa and compound 1 revealed several negative surface patches surrounding the active site cleft (Figure 2A). One

of these negative potentials spreads along the top rim of the active site starting from the area of the 99 and 60 loops, where side chains of 4 acidic residues (*Glu92*, *Asp60D*, *Glu62*, and *Asp63*) are exposed to the solvent, toward the 110 loop. Another notable surface patch of negatively charged side chains (*Glu146*, *Glu149*, and *Glu150*) protrudes along the activation loop below the active site cleft and close to the 186 loop (Figure 2A). These acidic residues (except *Asp60D*, which exists in both β-FXIIa and uPA) are not conserved among trypsin-like proteases and, therefore, the distribution of negatively charged residues along the FXIIa protease domain surface is unique for this enzyme. For example, in thrombin, the former patch is quite positive where *Arg60B*, *Arg86*, and *Arg89*
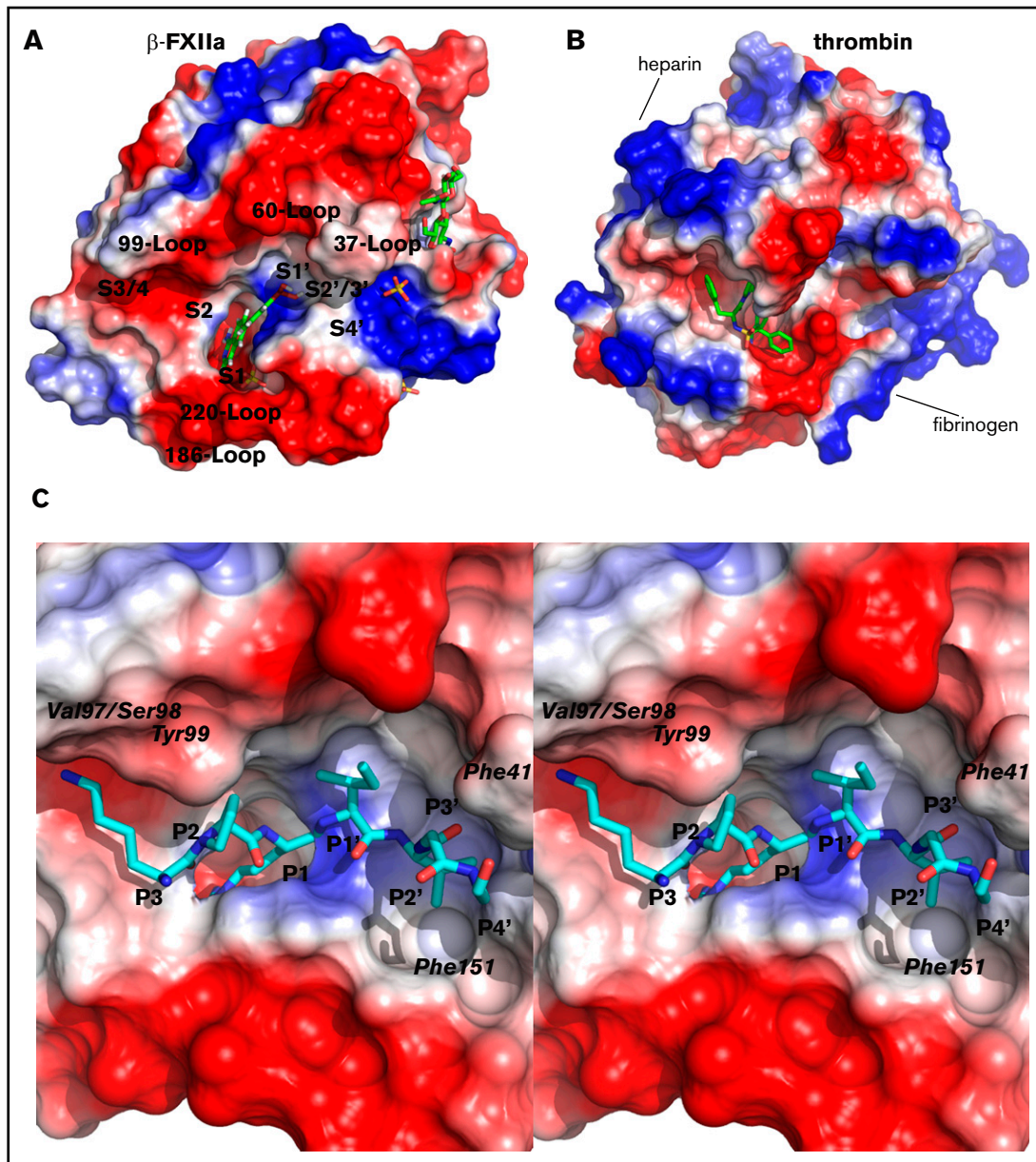
Figure 2. Surface charge of the β-FXIIa protease domain. (A) Solid surface colored according to the electrostatic potential calculated from the crystal structure of β-FXIIa in complex with compound 1. Positive surface is shown in blue and negative is in red. Some surface loops surrounding the active site cleft and positions of the substrate-binding sites (S4-S4′) are labeled. Noncovalent inhibitors, sugar residues, and sulfate ions are shown as sticks. (B) Thrombin surface charge was calculated from the crystal structure (PDB code 4UFD) of the enzyme in complex with the benzamidine-containing moiety. The inhibitor is shown in green sticks and the view of the thrombin molecule is that of the β-FXIIa molecule orientation shown in panel A. (C) Close-up stereo view of the FXIIa substrate from P3 to P4′ (ball and stick in cyan) interacting with the surface of the substrate-binding area. The residues P3 to P4′ modeled into the active site (see "Materials and methods") according to the canonical conformation are Lys-Pro-Arg-Ile-Val-Gly-Gly from the FXI amino acid sequence (residues 385-391; the scissile peptide bond is between the underlined residues). The positions of several FXIIa catalytic domain residues, which can form hypothetical interactions with the substrate in the encounter complex are indicated.

are surrounded by other alkaline residues, *Arg107*, *Arg129*, *Arg167*, and *Arg239*, to form the heparin-binding site (Figure 2B).[22] Also, the fibrinogen-binding site of the thrombin catalytic domain, which is positively charged, spatially overlays the latter negative site of the FXIIa protease domain.

Surface loops surrounding the active site of trypsin-like serine proteases are well known to be involved in many specific interactions with their cognate inhibitors, substrates, and cofactors.[22-24]

All main-chain atoms and most side chains of the surface loops are well defined by electron density in the crystal structures of human plasma β-FXIIa in complex with inhibitors. One of these loops, the β-FXIIa 37-surface loop, although having the same length of polypeptide segment as in trypsin and HGFA, adopts a quite different conformation and does not make any contact with the 60 loop, unlike hydrophobic interactions between side chains of *Ile35* and *Phe59* in HGFA or salt bridges between side chains of *Lys60*

and *Ser34/Tyr39* in trypsin. Simple modeling revealed 1 possible interaction between the carbonyl atom of *Phe41* of β-FXIIa and its substrate at the P3′ site (Figure 2C), which is also observed in the complex between HGFA and the inhibitory domain KD1 from inhibitor-1B.[25]

The 60-insertion loop resides directly adjacent to the 37 loop (Figures 1A and 2A) and has the same length as in HGFA, tPA, and uPA, is 4 residues longer than in trypsin and trypsinogen, and is 5 residues shorter than in thrombin (supplemental Figure 1). Despite the difference in the sequence, the 60 loop in FXIIa exhibits a main-chain conformation similar to the corresponding loops in HGFA, tPA, and uPA.

The 99 loop of β-FXIIa has a β-hairpin conformation like other related serine proteases, and bears the greatest resemblance to the corresponding loops with the same length in trypsin and tPA. However, this loop is more hydrophobic than in trypsin and tPA and, at its most exposed part, there are *Ser95-Pro96-Val97* residues in FXIIa instead of *Asn95-Ser96-Asn97* in trypsin or *Asp95-Asp96-Asp97* in tPA. The side chain of *Phe99* points toward the active site, and similar to trypsin and FXa, this residue could partially block the S2 subsite.[26,27] Based on simple modeling using the structure noncovalent complex between active HGFA and the KD1 domain from its cognate HAI-1 inhibitor, the P3 Lys of the FXIIa substrate (in this case, we used the FXI sequence from P3 to P4′; Table 2) projects toward the 99 loop of β-FXIIa and its side chain could form hydrogen bonds with the carbonyl atoms of *Val97* and *Ser98* (Figure 2C).

All atoms of the 140-autolysis loop, residues *143-152* except the side-chain atoms of *Glu150*, are well defined in the final $2F_o$-F_c density map. As in HGFA and t-PA, this loop is exposed to the solvent, and its main- and side-chain atoms have a relatively high average temperature factor of 88 Å^2 (the mean isotropic B factor for all atoms of the protein molecule is 58.0 Å^2 for the compound 1–β-FXIIa complex; Table 1), suggesting that the 3 acidic residues *Glu146*, *Glu149*, and *Glu150* might be involved in specific interactions with substrates, inhibitors, and cofactors. The side chain of *Tyr151*, similar to HGFA, trypsin, and t-PA, protrudes into the S2′ site (Figure 2C).

The 186 loop (*Gly184-Gly188*) of FXIIa is close to the entrance frame of the S1 pocket (*Gly221-Pro225*) and has the same length as in HGFA, trypsin, and uPA, 2 residues longer than in chymotrypsinogen, and 6 residues shorter than in tPA (supplemental Figure 1). The conformations of the loop main-chain residues *184-188* in human FXIIa, HGFA, and bovine trypsin are well superimposed within the RMSD value of 1.0 Å for their Cα residues. At the C terminus of the loop, there are main-chain hydrogen bonds between *Asp189-Ala190* and *Val16-Val17* located at the N terminus of the β-FXIIa protease domain.

There is a deletion at position *218*, and an insertion behind *Cys220* (*Asp221A*) in the β-FXIIa molecule, in common with trypsin, tPA, and HGFA, so this surface *210-220* segment is exposed to the solvent and is part of the top edge of the active site (Figure 1B). In contrast to trypsin, tPA and HGFA, but similar to coagulation factor Xa and thrombin, there is an insertion of 3 residues (*205A-205B-205C*) behind residue *205*; this results in a bulged surface loop located opposite to the active site and in salt bridges between this loop and a few residues of the heavy chain remnant.

Although β-FXIIa exists as a monomer in solution under physiological conditions, extensive intermolecular contacts between human

Table 2. Sequences around the scissile peptide bonds of the β-FXIIa substrates FXI and kallikrein and of the β-FXIIa endogenous inhibitors serpin C1, antithrombin, and PAI-1

	P4	P3	P2	P1	P1′	P2′	P3′	P4′	P5′
FXI	Ile	Lys	Pro	Arg	Ile	Val	Gly	Gly	Thr
Plasma kallikrein	Thr	Ser	Thr	Arg	Ile	Val	Gly	Gly	Thr
Serpin C1	Ser	Val	Ala	Arg	Thr	Leu	Leu	Val	Phe
Antithrombin	Ile	Ala	Gly	Arg	Ser	Leu	Asn	Pro	Asn
PAI-1	Val	Ser	Ala	Arg	Met	Ala	Pro	Glu	Glu

FXIIa protease domains in the crystal packing were observed (Figure 3). However, there is no biological or biochemical evidence that dimer or multimer formations have potential relevance for FXIIa functions.

Inhibitor binding

The β-FXIIa specificity pocket bordered by segments *Ile213-Cys220*, *Asp189-Ser195*, *Pro225-Thr229*, and disulfide bond *Cys191-Cys220* is practically identical to that of other active trypsin-like serine proteases. Two different small compound inhibitors in the crystal complex structures solved in this work are clearly defined by electron density in the enzyme primary site, with *Asp189* positioned at the bottom of the pocket to allow salt bridge formation with the basic group of the P1 moiety. The basic benzamidine molecule in the β-FXIIa–benzamidine complex is sandwiched between main-chain segments *Trp215-Gly216* and *Cys191-Gln192*, which is very similar, for instance, to the corresponding trypsin complex (PDB code 3MFJ), except for a 22° rotation around its long molecular axis (Figure 4A).[27] The amidinium group of the inhibitor makes a symmetric salt bridge with *Asp189*, also forming 2 salt bridges to the carbonyl oxygens of *Gly219* and *Ala190*.

Compound 1 binds to the FXIIa primary site in an extended conformation and its aminoisoquinoline group interacts productively with *Asp189* at the bottom of the S1 pocket (distances of 2.9 Å and 3.5 Å between 2 carboxylate oxygens and the amino group; Figure 4B). In addition, the other electrostatic interactions in the S1 pocket include the hydrogen bonds between the carbonyl oxygens of *Gly219* and *Ala190* and the amino group of the isoquinolone, but the *Gly219-Cys220* peptide bond flips by 60° compared with the benzamidine structure (Figure 4A). At the top of the enzyme primary site, compound 1 displaces the sulfate ion found in the benzamidine structure; the acid group of the inhibitor forms 2 salt bridges with the main-chain nitrogens of *Gly193* and *Asp194*. Continuous electron density links the inhibitor boron atom to the γ-O of *Ser195*, indicating the presence of a covalent linkage between the inhibitor and the enzyme.

Discussion

The discovery that both the zymogen and the activated form of FXII play significant physiological functions in vivo, such as the FXII growth factor activity or its essential contribution to thrombosis demonstrated in the f12 knockout mouse model, has generated great interest in the protein because it makes this coagulation factor an attractive target for new anticoagulants.[6-9] However, structural data for the activated form of FXII have been unavailable. In this study, we present the first structural characterization of human
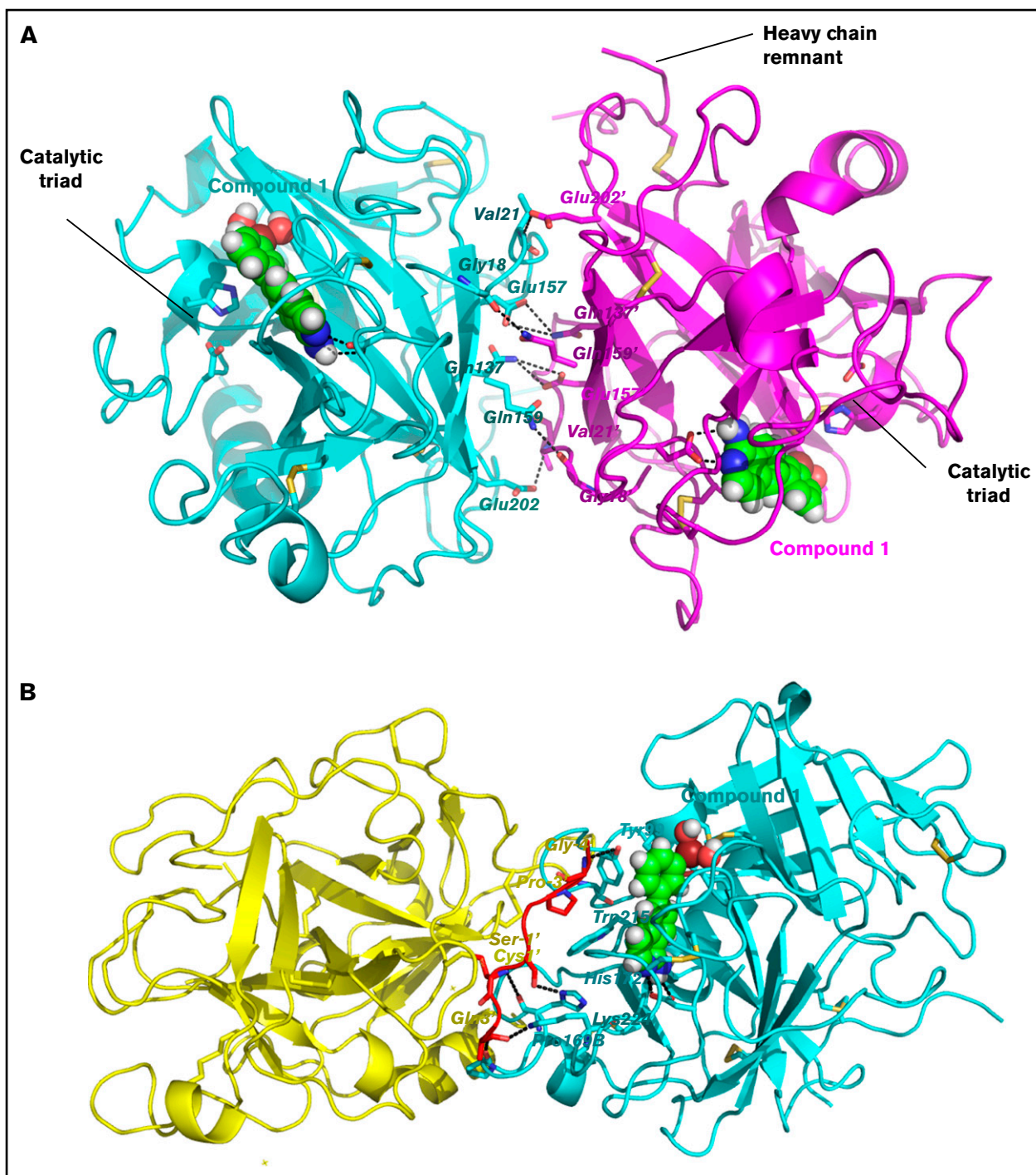
Figure 3. Two distinct intermolecular contacts between symmetry related molecules of β-FXIIa protease domain in the crystal packing. (A) First extensive interface. Ribbon representation of 2 protease domains shown in cyan and magenta. Contacting residues are labeled and shown as sticks. The residues of the neighboring molecule are denoted by the "'" mark. Salt bridges and hydrogen bonds are depicted by black dotted lines. Compound 1, buried in the active site, is shown as balls. Positions of the catalytic triad for both molecules and the heavy chain remnant for one molecule (on the right) are indicated. (B) Second contact region. The same as in panel A except the heavy chain remnant of the β-FXIIa molecule on the left (yellow cartoon) is shown in red.

plasma β-FXIIa in complex with 2 different inhibitors: benzamidine, the canonical basic inhibitor of serine proteases, and a small synthetic inhibitor that contains a boronic acid. The primary specificity site of β-FXIIa is practically identical to that of HGFA,

tPA, FXa, and thrombin, and there are only 2 differences in the FXIIa primary pocket compared with uPA, chymotrypsin, and trypsin where residues *Ala190* and *Ile213* are replaced by Ser and Val, respectively (supplemental Figure 1). Therefore, like HGFA,
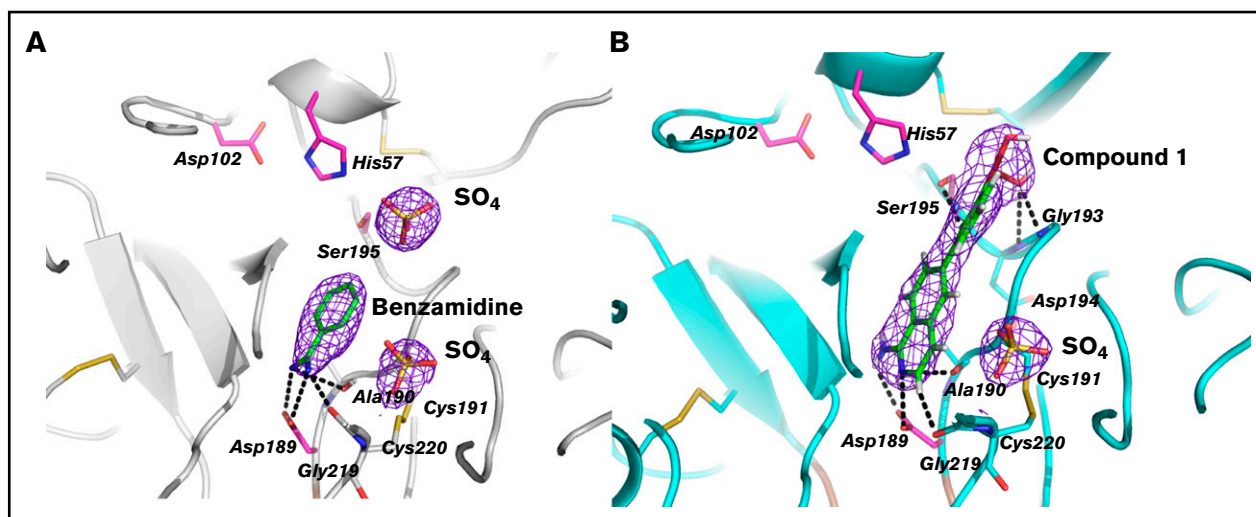
Figure 4. Binding interactions of small molecule inhibitors to β-FXIIa. (A) Close-up view of the inhibitor binding area in the benzamidine complex (gray ribbon). The benzamidine molecule (green sticks) bound to the S1 site, the residues defining the interaction with the inhibitors, the residues of the catalytic triad, and the bound phosphate ions (yellow) are shown as sticks. The electron density maps (blue) around the bound inhibitor and the ion are countered at 1σ. (B) The compound 1–β-FXIIa complex. The same view as in panel A.

thrombin, tPA, and the other related coagulation protease FXa, the primary pocket of β-FXIIa is slightly less polar. This could explain the FXIIa preference for substrate and cognate inhibitors with P1 Arg, as had been previously observed.[28,29] Another structural feature of β-FXIIa includes the restriction of the S2 subsite by the imposing side chain of *Tyr99* similar to thrombin and FXa (Figure 2C), explaining the preference of β-FXIIa for small residues at the P2. This argument is in agreement with the solution-phase fluorogenic peptide microarrays and single substrate studies that report the preference for Thr, Ser, and Gly in the P2 position, Met and Gln in the P3, and disfavor Glu at the P3.[30,31] These data are also consistent with the physiological substrates of FXIIa, FXI zymogen, and PPK, where the sequences around the scissile bonds are Lys-Pro-<u>Arg-Ile</u> and Ser-Thr-<u>Arg-Ile</u>, respectively (Table 2).

It is assumed that serpins, inhibitors of serine proteases, interact with their cognate proteases via their reactive center loops (RCLs), which adopt a canonical conformation from P4 to P3′ and make extensive contacts from both backbone and side chains to residues in proteases from P5 to P6′.[32] With the exception of RCL interactions, there may be extensive interactions between serpin exosites and surface loops surrounding the enzyme active site in the encounter serpin-protease complexes.[33] Serpin C1 esterase inhibitor is cognate inhibitor of β-FXIIa, and accounts for 92% of FXIIa inhibition in plasma by a serpin.[10] There are several structural models available for this serpin where the RCL is not ordered, making these C1INH structures unusable for modeling. PAI-1 is a trivial inhibitor of activated FXII.[34] Therefore, it could be possible to obtain a reliable structural model for polypeptide substrate-binding interactions based on a structural comparison between the crystal structure of β-FXIIa solved in this work and the available tPA–PAI-1 encounter complex structure.[35] Our simple modeling results provide a suitable model for this substrate interaction geometry of β-FXIIa (supplemental Figure 3A-B). There are also interactions between serpin exosites and protease surface loops; side chains of Thr205, Lys207, and Arg271 of PAI-1 might form salt bridges with *Gln192*, *Gln60*, and

Asp60A of β-FXIIa, respectively, similar to those in the tPA-PAI-1 encounter complex (supplemental Figure 3B).[35]

Figure 5 represents optimal superposition of the crystal structure of β-FXIIa onto that of the FXIIa zymogen[13] and can provide some insight into an examination of the FXII activation mechanism. The structure of the apo-form (nonbound with any ligands) of human FXII light chain (FXIIc, PDB code 4XDE, *Val16-Ser244*) lacking 9
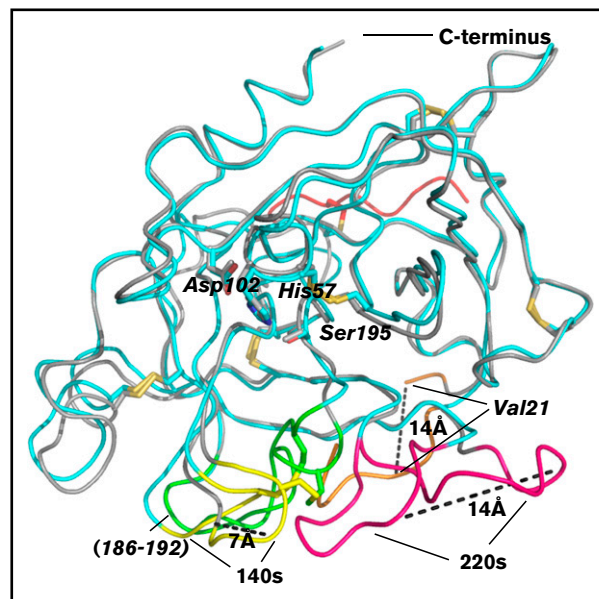
Figure 5. Structural comparison of the catalytic domains of β-FXIIa and FXIIc. Superposition of the β-FXIIa–compound 1 (cyan) reported in this study and the zymogen form of the FXII catalytic domain (FXIIc, residues *16-244*; PDB code 4XDE, gray cartoon). The greatest conformational changes between 2 structures are indicated by dashed lines.

residues of the heavy chain remnant revealed a typical zymogen conformation for the protease.[13] The largest conformational differences between 2 catalytic domains are found in the autolysis loop (residues *16-26*), in the activation domain (residues *186-192*), and in the adjacent surface 140 and 220 loops; in particular, the first ordered N-terminal residue *Val21* in the zymogen structure is 14 Å away from its position in the activated enzyme, the *Val16-Leu20* segment is completely missing in the zymogen, which results in the lack of an oxyanion hole, whereas the largest differences between the corresponding C-α atoms are 7 Å and 14 Å for the residues in the 140 and 220 loops, respectively (Figure 5). The main-chain conformations of other regions and the residues of the catalytic triad superimpose well with the RMSD value of 0.9 Å for 207 C-α atoms of the catalytic domains, excluding the above-mentioned labile regions.

Incompetent conformation of the protease domain of serine proteases was observed for the apo-form of HGFA expressed as a recombinant protein with 2 chains covalently linked together by a disulfide bond, whereas the same 2-chain construct in complex with the Kunitz domain inhibitor was expressed as an active enzyme.[25] Such conformational plasticity is also observed for other proteases, including thrombin activity regulation by sodium atoms and partial transformation of the inactive trypsinogen into a catalytically active enzyme in the presence of Ile-Val dipeptides without activation cleavage or conformation transition observed for apo-prostatin and prostatin.[36,37]

Here, we assume that the zymogen conformation of FXIIc could be explained by such conformational diversity. The presence of the heavy chain remnant in the protein construct seems not to be one of the main requirements for expression of the active form of FXII because other serine proteases, such as trypsin, kallikrein, and FXIa, lacking the short peptide segments covalently linked with the catalytic domain, have been crystallized in the canonical conformation.[20,37,38] Moreover, as was discussed earlier in this section, the HGFA construct having the remnant segment has been crystallized in both competent (in complex with an inhibitor) and incompetent (the apo-enzyme with no inhibitor bound) forms.[25] Also, the crystal structure of β-FXIIa, where the remnant chain does not make any interactions with the autolysis loop to stabilize the active conformation, supports this argument. Furthermore, in the zymogen-like crystal structure of FXIIc determined by Pathak et al,[13] an interaction between side chains of *Arg73* and *Asp194* was observed. The *Asp194* is well known to play a critical role in switching between the activated and inactive forms of trypsin-like serine proteases. Our structural study of β-FXIIa in complex with different inhibitors revealed the canonical interaction between the *Asp194* side chain and the N-terminal *Val16*.

In conclusion, a high-quality structure of human plasma β-FXIIa, which likely contributes to different pathological diseases via its biological functions, is presented. The new structural features of the active site cleft, substrate-binding sites, and the surface loop conformations of β-FXIIa may be a foundation for an efficient structure-based drug design platform targeting the FXIIa-driven plasma contact activities.

Acknowledgment

Authorship

Contribution: A.D. designed the crystallization experiments, analyzed data, solved crystal structures, and wrote the manuscript; A.S., Z.L., and H.S. conducted biochemical experiments; C.Y. synthesized compound 1; M.T.F. was responsible for oversight of the project; J.R.P. analyzed experimental data and wrote the manuscript; and all authors read and approved the final manuscript.

Conflict-of-interest disclosure: A.D. and M.T.F. are Shamrock Structures, LLC employees. A.S., C.Y., Z.L., H.S. and J.R.P. are Global Blood Therapeutics employees.

Correspondence: Alexey Dementiev, Shamrock Structures, 1440 Davey Rd, Woodridge, IL 60517; e-mail: adementiev@ shamrockstructures.com; and James R. Partridge, GBT, 400 East Jamie Ct, Suite 101, South San Francisco, CA 94080; e-mail: jpartridge@globalbloodtx.com.

References

1. Colman RW, Schmaier AH. Contact system: a vascular biology modulator with anticoagulant, profibrinolytic, antiadhesive, and proinflammatory attributes. *Blood*. 1997;90(10):3819-3843.

2. Samuel M, Pixley RA, Villanueva MA, Colman RW, Villanueva GB. Human factor XII (Hageman factor) autoactivation by dextran sulfate. Circular dichroism, fluorescence, and ultraviolet difference spectroscopic studies. *J Biol Chem*. 1992;267(27):19691-19697.

3. Colman RW. Contact activation (kallikrein-kinin) pathway: multiple physiologic and pathophysiologic activities. In: Colman RW, Mader VJ, Clowes AW, et al, eds. Hemostasis and Thrombosis: Basic Principles and Clinical Practice. 5th ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2006:107-130.

4. Müller F, Mutch NJ, Schenk WA, et al. Platelet polyphosphates are proinflammatory and procoagulant mediators in vivo. *Cell*. 2009;139(6):1143-1156.

5. Gailani D, Renné T. Intrinsic pathway of coagulation and arterial thrombosis. *Arterioscler Thromb Vasc Biol*. 2007;27(12):2507-2513.

6. Lämmle B, Wuillemin WA, Huber I, et al. Thromboembolism and bleeding tendency in congenital factor XII deficiency–a study on 74 subjects from 14 Swiss families. *Thromb Haemost*. 1991;65(2):117-121.

7. Ratnoff OD, Colopy JE. A familial hemorrhagic trait associated with a deficiency of a clot-promoting fraction of plasma. *J Clin Invest*. 1955;34(4): 602-613.

8. Leeb-Lundberg LM, Marceau F, Müller-Esterl W, Pettibone DJ, Zuraw BL. International union of pharmacology. XLV. Classification of the kinin receptor family: from molecular mechanisms to pathophysiological consequences. *Pharmacol Rev*. 2005;57(1):27-77.

9. Müller F, Renné T. Novel roles for factor XII-driven plasma contact activation system. *Curr Opin Hematol*. 2008;15(5):516-521.

10. Revak SD, Cochrane CG. The relationship of structure and function in human Hageman factor. The association of enzymatic and binding activities with separate regions of the molecule. *J Clin Invest*. 1976;57(4):852-860.

11. Schapira M. Major inhibitors of the contact phase coagulation factors. *Semin Thromb Hemost*. 1987;13(1):69-78.

12. Weitz JI. Factor Xa and thrombin as targets for new oral anticoagulants. *Thromb Res*. 2011;127(suppl 2):S5-S12.

13. Pathak M, Wilmann P, Awford J, et al. Coagulation factor XII protease domain crystal structure. *J Thromb Haemost*. 2015;13(4):580-591.

14. Cool DE, Edgell CJ, Louie GV, Zoller MJ, Brayer GD, MacGillivray RT. Characterization of human blood coagulation factor XII cDNA. Prediction of the primary structure of factor XII and the tertiary structure of beta-factor XIIa. *J Biol Chem*. 1985;260(25):13666-13676.

15. Otwinowski Z, Minor W. Processing of x-ray diffraction data collected in oscillation mode. *Methods Enzymol*. 1997;276:307-326.

16. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. *J Appl Cryst*. 2007;40(Pt 4):658-674.

17. Winn MD, Ballard CC, Cowtan KD, et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr*. 2011;67(Pt 4): 235-242.

18. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr*. 2010;66(Pt 4):486-501.

19. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr*. 1997;53(Pt 3):240-255.

20. Tang J, Yu CL, Williams SR, et al. Expression, crystallization, and three-dimensional structure of the catalytic domain of human plasma kallikrein. *J Biol Chem*. 2005;280(49):41077-41089.

21. Navaneetham D, Jin L, Pandey P, et al. Structural and mutational analyses of the molecular interactions between the catalytic domain of factor XIa and the Kunitz protease inhibitor domain of protease nexin 2. *J Biol Chem*. 2005;280(43):36165-36175.

22. Bode W, Turk D, Karshikov A. The refined 1.9-A x-ray crystal structure of D-Phe-Pro-Arg chloromethylketone-inhibited human α-thrombin: structure analysis, overall structure, electrostatic properties, detailed active-site geometry, and structure-function relationships. *Protein Sci*. 1992;1(4):426-471.

23. Horrevoets AJG, Tans G, Smilde AE, van Zonneveld AJ, Pannekoek H. Thrombin-variable region 1 (VR1). Evidence for the dominant contribution of VR1 of serine proteases to their interaction with plasminogen activator inhibitor 1. *J Biol Chem*. 1993;268(2):779-782.

24. Madison EL, Goldsmith EJ, Gerard RD, Gething MJH, Sambrook JF. Serpin-resistant mutants of human tissue-type plasminogen activator. *Nature*. 1989; 339(6227):721-724.

25. Shia S, Stamos J, Kirchhofer D, et al. Conformational lability in serine protease active sites: structures of hepatocyte growth factor activator (HGFA) alone and with the inhibitory domain from HGFA inhibitor-1B. *J Mol Biol*. 2005;346(5):1335-1349.

26. Nar H, Bauer M, Schmid A, et al. Structural basis for inhibition promiscuity of dual specific thrombin and factor Xa blood coagulation inhibitors. *Structure*. 2001;9(1):29-37.

27. Bode W, Schwager P. The refined crystal structure of bovine beta-trypsin at 1.8 Å resolution. II. Crystallographic refinement, calcium binding site, benzamidine binding site and active site at pH 7.0. *J Mol Biol*. 1975;98(4):693-717.

28. Kato H, Adachi N, Ohno Y, Iwanaga S, Takada K, Sakakibara S. New fluorogenic peptide substrates for plasmin. *J Biochem*. 1980;88(1):183-190.

29. McRae BJ, Kurachi K, Heimark RL, Fujikawa K, Davie EW, Powers JC. Mapping the active sites of bovine thrombin, factor IXa, factor Xa, factor XIa, factor XIIa, plasma kallikrein, and trypsin with amino acid and peptide thioesters: development of new sensitive substrates. *Biochemistry*. 1981;20(25): 7196-7206.

30. Cho K, Tanaka T, Cook RR, et al. Active-site mapping of bovine and human blood coagulation serine proteases using synthetic peptide 4-nitroanilide and thio ester substrates. *Biochemistry*. 1984;23(4):644-650.

31. Gosalia DN, Salisbury CM, Ellman JA, Diamond SL. High throughput substrate specificity profiling of serine and cysteine proteases using solution-phase fluorogenic peptide microarrays. *Mol Cell Proteomics*. 2005;4(5):626-636.

32. Dementiev A, Simonovic M, Volz K, Gettins PGW. Canonical inhibitor-like interactions explain reactivity of α1-proteinase inhibitor Pittsburgh and antithrombin with proteinases. *J Biol Chem*. 2003;278(39):37881-37887.

33. Dementiev A, Petitou M, Herbert JM, Gettins PGW. The ternary complex of antithrombin-anhydrothrombin-heparin reveals the basis of inhibitor specificity. *Nat Struct Mol Biol*. 2004;11(9):863-867.

34. Hedner U, Martinsson G. Inhibition of activated Hageman factor (factor XIIa) by an inhibitor of the plasminogen activation (PA inhibitor). *Thromb Res*. 1978;12(6):1015-1023.

35. Gong L, Liu M, Zeng T, et al. Crystal structure of the Michaelis complex between tissue-type plasminogen activator and plasminogen activators inhibitor-1. *J Biol Chem*. 2015;290(43):25795-25804.

36. Spraggon G, Hornsby M, Shipway A, et al. Active site conformational changes of prostasin provide a new mechanism of protease regulation by divalent cations. *Protein Sci*. 2009;18(5):1081-1094.

37. Bode W, Schwager P, Huber R. The transition of bovine trypsinogen to a trypsin-like state upon strong ligand binding. The refined crystal structures of the bovine trypsinogen-pancreatic trypsin inhibitor complex and of its ternary complex with Ile-Val at 1.9 Å resolution. *J Mol Biol*. 1978;118(1):99-112.

38. Jin L, Pandey P, Babine RE, et al. Crystal structures of the FXIa catalytic domain in complex with ecotin mutants reveal substrate-like interactions. *J Biol Chem*. 2005;280(6):4704-4712.