manuscript gi5019 for review

Crystal structures of the recombinant β-Factor XIIa protease with bound Thr-Arg and Pro-Arg substrate mimetics

monika pathak*, rosa manna, chan li, bubacarr kaira, Badraldin Kareem, Benny Belviso, camila Bonturi, ingrid dreveny, peter fischer, Lodewijk dekker, Maria oliva and jonas emsley

Category: *research papers*

Co-editor:

Professor R. Garratt

Telephone: +55 16 3373 9874/9846

Fax: +55 16 3373 9881

Email: richard@ifsc.usp.br

Contact author:

jonas emsley

School of Pharmacy, university of nottingham, university of nottingham, nottingham, Nottingham, ng72rd, United Kingdom

Telephone: 01158467092

Fax: 01158468002

Email: jonas.emsley@nottingham.ac.uk

Crystal structures of the recombinant β-Factor XIIa protease with bound Thr-Arg and Pro-Arg substrate mimetics

Monika Pathak[1]*, Rosa Manna[1]*, Chan Li[1], Bubacarr G Kaira[1], Badraldin Kareem Hamad[1], Benny Danilo Belviso[1], Camila R Bonturi[2], Ingrid Dreveny[1], Peter M Fischer[1], Lodewijk V Dekker[1], Maria Luiza Viela Oliva[2], and Jonas Emsley[1¶].

[1] Centre for Biomolecular Sciences, School of Pharmacy, University of Nottingham, UK.

[2] Biochemistry Department, Federal University of São Paulo, 04044-020, São Paulo - SP, Brazil

¶To whom correspondence should be addressed: Jonas Emsley, School of Pharmacy, Centre for Biomolecular Sciences, University of Nottingham, Nottingham NG7 2RD, United Kingdom; jonas.emsley@nottingham.ac.uk; Tel.(+44) 115-8467092; Fax. (+44) 115- 8468002

**joint first authors.*

Synopsis: Coagulation factor XII (FXII) is a key initiator of the contact pathway and kinin generation. We report the first peptido-mimetic complex structure for the activated protease domain βFXIIa. These crystal structures of βFXIIa provide insight into serine protease substrate and inhibitor recognition.

Abstract

Coagulation factor XII (FXII) is a key initiator of the contact pathway which contributes to inflammatory pathways. FXII circulates as a zymogen, which when auto-activated forms factor XIIa (FXIIa). We report production of recombinant FXIIa protease domain (βFXIIa^{His}) with yields of ~1-2 mg/L of insect cell culture. A second construct utilised an N-terminal maltose-binding protein (MBP) fusion (MBP-βFXIIa^{His}). Crystal structures were determined for MBP-βFXIIa^{His} in complex with inhibitor D-Phe-Pro-Arg chloromethyl ketone (PPACK) and βFXIIa^{His} in isolation. The βFXIIa^{His} structure revealed the S2 and S1 pockets were occupied with Thr and Arg residues, respectively, from an adjacent molecule in the crystal. The Thr-Arg sequence mimics the P2-P1 FXIIa cleavage site residues present in the natural substrates prekallikrein and FXII and Pro-Arg (from PPACK) mimics the Factor XI cleavage site. A comparison of the βFXIIa^{His} structure with the available crystal structure of the zymogen-like FXII protease revealed large conformational changes centred around the S1 pocket and an alternate conformation for the 99-loop, Tyr99 and the S2 pocket. Further comparison with activated protease structures for factors IXa and Xa, which also have residue Tyr99, reveals that a more open form of the S2 pocket only occurs in the presence of a substrate mimetic. FXIIa inhibitors EctTI and Infestin-4 have Pro-Arg and Phe-Arg P2-P1 sequences respectively and we also describe the interactions these inhibitors form with βFXIIa. These structural studies of βFXIIa provide insight into substrate and inhibitor recognition and establish a scaffold for structure-guided drug design of novel anti-thrombotic and anti-inflammatory agents.

Keywords: coagulation factor XII; serine protease; crystal structure; inhibitor complex

1. Introduction

The contact activation system is initiated via auto-activation of the serine protease coagulation factor XII (FXII) (Weidmann *et al.*, 2017). Activated FXII (FXIIa) cleaves substrates coagulation factor XI (FXI), to initiate the intrinsic pathway of coagulation, and prekallikrein (PK), which results in the cleavage of kininogen and bradykinin (BK) formation (Maas & Renne, 2018). A number of studies have established that inhibition of FXII activity reduces the formation of pathological thrombi without compromising physiological haemostasis in both murine and primate models of cardiovascular disease (Renne *et al.*, 2012, Kleinschnitz *et al.*, 2006, Matafonov *et al.*, 2014). FXII has also been implicated in inflammatory pathways (Hess *et al.*, 2017, Bender *et al.*, 2017) and hereditary angioedema via identification of a gain-of-function mutation in the *F12* gene (encodes FXII), which results in excessive formation of BK (Cichon *et al.*, 2006).

Natural inhibitors of the FXII protease in humans include the serpin C1INH (Nickel *et al.*, 2017) and histidine rich glycoprotein (HRG) (MacQuarrie *et al.*, 2011). FXII inhibitors from exogenous sources have been described including corn trypsin inhibitor (CTI) (Mahoney *et al.*, 1984), *Enterolobium contortisiliquum* trypsin inhibitor (EcTI) (Batista *et al.*, 1996), *Triatoma infestans* infestin-4 (Campos *et al.*, 2004) and *Escherichia coli* ecotin (Ulmer *et al.*, 1995). CTI is widely used as an inhibitor that is selective for FXIIa and the intrinsic pathway of coagulation, but does not affect the extrinsic pathway. *In vivo* studies in animal models of thrombosis demonstrate infestin-4 provides protection from myocardial infarction and ischemic stroke without affecting hemostasis (Krupka *et al.*, 2016). The monoclonal antibody 3F7 targeting the FXIIa protease domain was shown to block artificial surface-induced thrombosis in mice and rabbits (Worm *et al.*, 2015, Larsson *et al.*, 2014).

The principal protease that activates FXII is plasma kallikrein (PKa). FXII has a major cleavage site between residues Arg353 and Val354, generating αFXIIa and additional cleavage sites are shown in Fig. 1. Auto-activation of FXII via contact with anionic surfaces

in the presence of Zn^{2+} ions is a second route to forming αFXIIa (Ivanov *et al.*, 2017). αFXIIa consists of two chains, the N-terminal heavy chain (353 residues, molecular weight (MW) ~50 kDa) and the C-terminal light chain (243 residues, MW ~28 kDa) connected via a disulfide bridge between Cys340 and Cys467 (Fig. 1). A further cleavage of αFXIIa between residues Arg334-Asn335 produces β-factor XII (βFXIIa), which includes the light chain and nine residues of the heavy chain, termed the heavy chain remnant (HCR) (Pathak *et al.*, 2015).

To characterise enzyme kinetics and perform protein crystallization experiments milligram quantities of purified βFXIIa protease are required. The insect cell-based Drosophila expression system (DES, Invitrogen) is a preferred choice for heterologous protein expression due to high yields and eukaryotic post-translational modifications. Trypsin-like serine proteases have a high degree of flexibility in their loop structures, which is known to prevent crystallization. One approach to stabilize flexible proteins is described by Moon *et al.* 2010, utilising maltose binding protein (MBP) with surface entropy reduction (SER) mutations as an N-terminal carrier protein to enhance protein crystallization (Moon *et al.*, 2010). MBP has been reported to enhance the solubility of proteins when expressed as an N-terminal fusion in *E. coli* (Waugh, 2016, Jin *et al.*, 2017) and mammalian expression systems (Reuten *et al.*, 2016, Bokhove *et al.*, 2016) but to date this has not been reported for insect cell expression systems. We describe the expression of the βFXIIa protease domain as a secreted MBP fusion using DES, which facilitated the crystal structure determination for MBP-βFXIIa^{His} in the active conformation in complex with peptidomimetic inhibitor D-Phe-Pro-Arg chloromethylketone (PPACK). Another construct encoding recombinant βFXIIa^{His} without the N-terminal fusion was also purified successfully and crystallized in the absence of inhibitor, stabilized by a crystal contact that mimics P1 (Arg) and P2 (Thr) residues from the activation loop of natural substrates FXII and PK. The latter βFXIIa^{His} structure provides a rare example of a trapped protease-product complex and reveals a more open conformation of the S2 pocket than the MBP-βFXIIa^{His}-PPACK complex and the plasma purified βFXIIa-

benzamidine complex crystal structure (Fig. S1). We also utilised the βFXIIa^{His} crystal structure to examine the interactions formed with EcTI a kunitz type trypsin inhibitor and Infestin-4 a kazal-type inhibitor.

2. Materials and methods

All chemical and reagents were purchased from Sigma–Aldrich (Dorset, UK) unless otherwise stated and the purification columns used were purchased from GE Healthcare.

2.1 Cloning and expression of βFXIIa^{His} and MBP-βFXIIa^{His}

Based on the FXII mature protein sequence numbering, the gene fragment Asn335–Ser596 coding for human βFXIIa was cloned into a pMT-PURO (Addgene) expression plasmid alone or with a noncleavable N-terminal MBP tag (Moon *et al.*, 2010, Ullah *et al.*, 2008) for expression in the DES (Invitrogen, Thermo scientific Fischer, United States). To generate the MBP-βFXIIa^{His} construct, βFXIIa was fused to the MBP carrier protein with a five amino acid residues (AlaAlaAlaAlaSer) linker, as reported by Moon *et.al.*, 2010 using the In-Fusion Advantage PCR Cloning KIT protocol (Clontech Laboratories, United States) and then subsequently cloned into the pMT-PURO vector. PCR primers were designed in order to share 15 bases of homology with the sequence at the ends of the linearized vector, MBP or FXII. The following set of primer pairs: 5'CTCGCTCGGG*AGATCT*AAAATCGAAGAAGGTAAACTGGA3' and 5'CGGGCCCTTAAGGCTAGCTGCGGCCGCATTAGTCTG3');

5'GCGGCCGCAGCTAGCGAATTCGGCCCGCTGAGCTGC3' and 5'ATGACCGGT*ACGCGT*GCTCACGGTATGTTCGC3' were used for MBP and FXII, respectively. The PCR product encoding βFXIIa^{His}, was amplified using the primer pairs: 5'GTCG*AGATCT*CGGACCCCGCCTCAGTCC3';5'GCGC*ACGCGT*GGAAACGGTGTGC TCCCGGA3'; then digested with the appropriate restriction enzyme (BglII and MluI) and finally ligated into pMT-PURO between the BglII and MluI restriction sites (Quick Ligation

Kit from NEB). The pMT-PURO vector has an N–terminal *Drosophila* homologue of the immunoglobulin binding chaperone protein secretion signal (BiP) to enable secretion of recombinant proteins and a six His-tag sequence at the C-terminus (Iwaki & Castellino, 2008). Plasmid DNA was purified with the commercially available Sigma–Aldrich Miniprep Kit and ethanol precipitated prior to transfection into *D. melanogaster* S2 cells. The cells were maintained in Schneider's *Drosophila* Medium (GIBCO, Fisher Scientific UK Ltd) supplemented with 10% fetal bovine serum (Sigma) and transfected (2-3 x 10^6 cells/mL, >95% viability 5 mL in 25 cm^2 flask (Nunc, Germany)) using standard calcium phosphate transfection methods. Selection of pMT-PURO plasmid transfected cells was performed using 10 µg/ml of puromycin antibiotic. Stable cells expressing the recombinant MBP-βFXIIa^{His} were confirmed based on immunoreactivity with an anti-His tag monoclonal antibody using western blot (Towbin *et al.*, 1979). When cells reached a density of 9×10^6 cells/mL they were passaged by diluting 1:2 with Serum-free Express Five SFM insect culture medium (GIBCO, Fisher Scientific UK Ltd) supplemented with 20 mM L-glutamine and finally cultured in a tissue culture shaker flask (Corning Incorporated, Corning, NY, USA) for large-scale expression with ≤2% serum. Protein expression was induced by supplementing culture media with $CuSO_4$ to a final concentration of 0.5 mM. After 7 days at 28° C, culture media were harvested, centrifuged and the supernatant was used for purification of the recombinant protein.

2.2 Protein purification of βFXIIa^{His} and MBP-βFXIIa^{His}

Purification of recombinant βFXIIa^{His} was initially performed by Ni-Sepharose affinity (His-Excel column) chromatography, with equilibration buffer containing 50 mM HEPES, pH 8.3, 500 mM NaCl, 5 mM benzamidine and 10% glycerol, followed by a wash step with 20 mM imidazole and eluted in presence of 500 mM imidazole. The protein obtained from the Ni-affinity column was subject to anion exchange chromatography (Resource Q) with buffer A containing 20 mM Tris HCl, pH 8.3, 5 mM benzamidine and

10% glycerol, and buffer B containing buffer A with 1 M NaCl. Gel filtration chromatography (HiLoad 16/60 Superdex 75 16/60 pg) was used for the final purification step with buffer containing 50 mM Na-HEPES at pH 8.3, 500 mM NaCl, 5 mM benzamidine, 10% glycerol. MBP-βFXIIa^{His} was initially purified using anion exchange chromatography with a Capto Q column with buffer A containing 20 mM Tris HCl, pH 8.0 and buffer B containing 20 mM Tris HCl, pH 8.0, 1 M NaCl. The second step comprised Ni-sepharose affinity chromatography with buffer A containing 50 mM Tris HCl, pH 8.0, 200 mM NaCl, 20 mM imidazole and buffer B containing 50 mM Tris HCl at pH 8.0, 200 mM NaCl, 1 M imidazole. This construct was finally purified by gel filtration (HiLoad 16/60 Superdex 200 16/60 pg) chromatography in buffer 20 mM Tris HCl, pH 8.0, 200 mM NaCl, 40 mM D-(+)-maltose. The purified protein was concentrated in 20 mM Tris HCl pH 8.0, 50 mM NaCl and 5 mM D-(+)-maltose. The yield of MBP-βFXIIa^{His} and βFXIIa^{His} varied between 1.3-2.0 mg/L and 0.8-1.0 mg/L of culture media, respectively.

2.3 Assays of FXIIa activity

The enzymatic activity of βFXIIa was assessed by measuring the change in absorbance at 405 nm associated with p-nitroaniline (pNA), formed upon cleavage of the chromogenic substrate, S-2302 (Chromogenix, Epsom, UK). 150 nM βFXIIa was tested against increasing concentrations of S-2302 (0.4, 1, 2, 3, 4 and 6 mM). Commercial βFXIIa (Enzyme Research Laboratories, Inc.), MBP-βFXIIa^{His} and βFXIIa^{His} activity was assayed using 10 µL of protein added to 10 µL of substrate and diluted with 80 µL of phosphate-buffered saline (pH 7.4) at 37 °C. The change in OD_{405} was monitored with a ENVISION multimode plate reader. An average of four independent experiments was used to determine the enzyme kinetics. Each substrate concentration was measured in quadruplicate (Fig. S2). The kinetic parameters were calculated from the experimental data using Graphpad Prism with the non-linear Michaelis-Menten algorithm.

2.4. βFXIIa^{His} and MBP-βFXIIa^{His} crystallization and structure determination

The purified βFXIIa^{His} protein was concentrated to 14 mg/mL for crystallization trials using screens from Qiagen, Molecular Dimension and Hampton Research (Aliso Viejo, CA, USA) sitting drop plates. Crystals grew from conditions including 0.1 M Tris-HCl, pH 8.0 and 1.5 M $(NH_4)_2SO_4$ at 19 °C. MBP-βFXIIa^{His} was co-crystallized with the cholormethylketone based inhibitor, D-Phe-Pro-Arg-chloromethylketone (PPACK), which covalently modifies the Ser195 and His57 residues (Bode *et al.*, 1989). The MBP-βFXII^{His}-PPACK complex was prepared by adding a 10-fold excess of PPACK and incubating the sample at 4 °C overnight. The crystallization of the MBP-βFXIIa^{His}-PPACK complex was carried out using two different protein concentrations, 8 mg/mL and 6 mg/mL. Crystals of MBP-βFXIIa^{His}-PPACK suitable for data collection were grown in 0.1 M sodium citrate, pH 5.6 with 10% PEG 4000 and 0.15 M $MgCl_2$ at 19 °C. Crystals obtained for both MBP-βFXII^{His}-PPACK and βFXIIa^{His} were harvested, cryo-protected in the crystallization solution with 25% glycerol and flash cooled in liquid nitrogen. X-ray diffraction data were collected at the beamline I04 at Diamond Light Source, UK for both βFXIIa^{His} and MBP-βFXIIa^{His}-PPACK, which diffracted to resolutions of 2.54 Å and 4 Å, respectively. For MBP-βFXIIa^{His}-PPACK crystal, diffraction data were processed with XDS (Kabsch, 2010) then scaled and merged with AIMLESS (Evans & Murshudov, 2013). The structure was determined by molecular replacement using PHASER (McCoy *et al.*, 2007) from the CCP4 suite with search templates of the FXII protease domain crystal structure (PDB code: 4XDE) and the structure of MBP (PDB code: 1HSJ). For βFXIIa^{His} the diffraction data were processed with XIA2-DIALS (Evans, 2006, Winter *et al.*, 2018) and the structure was determined using molecular replacement with one molecule in the asymmetric unit. High quality electron density maps resulted and model building was performed with Coot (Emsley & Cowtan, 2004) and refined with REFMAC (Murshudov *et al.*, 2011) with crystallographic statistics presented in Table 1.

2.5 Interaction of βFXIIa^{His} and EcTI

2.5.1 Gel filtration

EcTI was purified as previously described (Batista *et al.*, 1996). βFXIIa^{His} was concentrated to 500 μM and EcTI was added to a molar ratio of 1:1 then analyzed on analytical gel filtration column (Superdex 200 Increase 10/300 GL) in 20 mM Tris HCl, pH 8.0 and 200 mM NaCl, 5% glycerol. The EcTI-βFXIIa^{His} complex peak fraction was analyzed by SDS-PAGE. Gel filtration standards (GE healthcare) were run to calibrate the column and calibration curve revealed a calculated molecular weight of the EcTI-βFXIIa^{His} complex peak of ~43 kDa.

2.5.2 SPR binding studies

SPR experiments were performed on a Biacore 3000 Instrument (BIAcore) at 25 °C as described previously (Wong *et al.*, 2016). EcTI was covalently coupled on a CM5 sensor chip using an amine coupling kit (GE healthcare) with N-hydroxysuccinimide and N-ethyl-N-(3-(dimethylamino)propyl) carbodiimide activation. EcTI was immobilized at approximately 300 response units (RU) and another flow cell was subjected to the identical immobilization procedure without EcTI protein for use as a control. The binding of βFXII to EcTI was studied by injecting varying concentrations of βFXII for 1200 seconds in a running buffer (20 mM HEPES buffer, pH 7.4, 150 mM NaCl, 50 μM EDTA, 0.005% polysorbate 20) at a flow rate of 50 μL/min. To assess any nonspecific binding, the analyte βFXIIa (both commercial βFXIIa and recombinant βFXIIa^{His}) was also injected over the control flow cell. Binding curves were analyzed on the basis of the SPR RU recorded for each analyte protein concentration and data were processed with an appropriate fitting model using the BIA evaluation 4.1 to measure association and dissociation rate constants (K_D, k_a and k_d).

2.5.3 Molecular docking

Molecular docking was performed using CLUSPRO (Kozakov *et al.*, 2017) as described previously (Hamad *et al.*, 2017) and utilized the βFXIIa^{His} crystal structure and the available crystal structure of EcTI (Protein Data Bank (PDB) code 4J2Y) and infestin-4 (PDB code 2ERW). Molecular graphics and figures were created using PyMOL.

3. Results

3.1 Two recombinant forms of the active FXII protease

Constructs for the MBP-βFXIIa^{His} fusion and isolated protease domain βFXIIa^{His} (Fig. 1) were expressed in Schneider 2 (S2) cells using the pMT-PURO vector and subsequently purified from culture media using affinity, ion exchange and size exclusion chromatography. SDS-PAGE gels under reducing conditions revealed the purified βFXIIa^{His} migrated as single band of ~30 kDa and MBP-βFXIIa^{His} as two bands of ~30 kDa and ~40 kDa, corresponding to βFXIIa and MBP, respectively, illustrating that during the purification the MBP-βFXIIa^{His} activation loop has been cleaved.

To characterize the catalytic activity of MBP-βFXIIa^{His}, the hydrolysis of chromogenic substrate S-2302 was monitored at 405 nm. The kinetic parameters for recombinant MBP-βFXIIa^{His}, αFXIIa and βFXIIa were calculated from the initial reaction rates as a function of the concentration of S-2302 and the Michaelis constant K_m for MBP-βFXIIa^{His} is comparable to commercial plasma purified βFXIIa and αFXIIa (Fig. S2; Table 1).

3.2 Crystal structure of βFXIIa^{His}

βFXIIa^{His} crystallized in the tetrahedral space group $P4_12_12$ and data were collected to 2.54 Å resolution with one molecule in the asymmetric unit. The crystal structure of βFXIIa^{His} spans residues Asn^{335}-Ser^{596} (native FXII sequence numbering) with high-quality electron density and crystallographic refinement statistics (Table 2, Fig. S3). Our cloning strategy using the MluI restriction site resulted in the recombinant βFXIIa^{His} having four extra residues at the C-terminus prior to the His-tag with sequence Thr^{1}-Arg^{2}-Thr^{3}-Gly^{4}-(6X)His (Fig. 1). We have followed a chymotrypsin-based numbering system throughout for FXII residues unless otherwise indicated and the C-terminal non-native amino acids are numbered in superscript (i.e. Thr^{1}) prior to the His-tag. Unexpectedly, Thr^{1}-Arg^{2} was observed in the electron density maps forming a crystal contact, whereas the remaining C-terminal residues Thr^{3}-Gly^{4}-(6X)His were absent (Fig. 2*A*, S3). The Arg^{2} side chain is observed in the S1 binding pocket and its guanidinium group forms the characteristic salt bridge to the Asp189 carboxylate (Fig. 2*B* and S3). The Thr^{1} side chain projects into the S2 pocket, formed by side chains of His57, Tyr99, Phe94, Pro96 and Gln60 and the Thr^{1} main chain carbonyl forms a hydrogen bond to the side chain of Gln192 (Fig. 2*B*). FXII residue Gly216 from the S2 pocket also forms two hydrogen bonds to the main chain of Ser244 from the crystal contact which are characteristic S2 pocket interactions for bound peptide mimetic P3 residues observed in other protease complex crystal structures (Kristensen *et al.*, 2016, Bode *et al.*, 1989). Thus the βFXIIa^{His} structure likely represents an autoproteolytic cleavage of the Arg^{2}-Thr^{3} peptide bond to produce a crystal contact that presents a product complex with the C-terminal Arg^{2} carboxylate forming electrostatic interactions with the βFXIIa^{His} oxyanion hole. A superposition of the recombinant βFXIIa^{His} structure with the plasma purified βFXIIa-benzamidine complex structure (Dementiev *et al.*, 2018) illustrates key differences arise in the area of the substrate selectivity S2 pocket (Fig. S1) as outlined below.

3.3 Comparison of βFXIIa^{His} with FXII zymogen protease structures

We have previously reported two zymogen-like crystal structures for the FXII protease spanning the protease domain alone (amino acids 354 to 596) (Pathak *et al.*, 2015). These are termed (i) FXIIc (PDB code: 4XDE), which corresponds to a construct where the N-terminus is blocked by addition of two extra amino acids Arg-Ser and has very low enzymatic activity and (ii) FXIIac, (PDB code: 4XE4) which has the native N-terminal Val354 residue and 10-fold higher enzyme activity compared to FXIIc but 1000-fold lower activity compared to βFXIIa. FXIIac is lacking the additional N-terminal residues of the HCR found in βFXIIa, which we speculated are required for efficient catalytic activity (Pathak *et al.*, 2015). Comparison of βFXIIa^{His} with FXIIc reveals the expected major differences in the conformation of the 140-loop (also termed the autolysis loop), 180-loop and 220-loop as is typical for trypsin-like serine proteases (Fig. 2*C*). In βFXIIa^{His}, the 140-loop forms interactions with the C-terminal β-barrel, which is in contrast to the 140-loop that interacts with the N-terminal β-barrel in both the FXIIc and FXIIac structures.

Additional changes not previously observed are the movement of the 99-loop away from the catalytic serine in βFXIIa^{His} and the Tyr99 sidechain has an alternate conformation compared to the FXIIc zymogen protease structure. The consequence is the Tyr99 sidechain moves away from the close packing interaction with His57 and thus opens up the S2 pocket in βFXIIa (Fig. 2*D,E*), which notably comes at the expense of the S3 pocket as the Tyr99 sidechain partially occludes this by packing against Trp215.

3.4 βFXIIa surface charge features

We have previously observed ridges of negative charge (termed R1 and R2) in the FXIIc structure surrounding the catalytic residues (Pathak *et al.*, 2015). A comparison of FXIIc with the active βFXIIa^{His} structure reveals a substantial re-arrangement of charge at the R2 ridge 140-loop residues Glu151B, Glu151E and Glu151F, which move from the C-terminal β-

barrel closer to the S1 pocket and contributes to the negative charge flanking the region of the S1' pocket together with residue Asp60A (Fig. 3*A* and *B*). The 140-loop is known to contribute to the interaction of proteases with serpins (Gong *et al.*, 2015, Baglin *et al.*, 2002) and thus the differences in this region observed between FXIIa and proteases such as FXIa and thrombin from the extrinsic pathway may contribute to the selectivity of inhibitor binding which is examined below.

Another distinctive feature observed in the zymogen-like FXIIc structure is the hydrophobic H1 pocket, which is positioned at the centre of the N-terminal β-barrel and opposite to the S1 pocket. In the βFXIIa^{His} structure the H1 pocket is closed due to the position of Trp35, which forms a lid on the pocket and this is likely to be the dominant conformation (Fig. 3*A*).

3.5 MBP-βFXIIa^{His}-PPACK crystal structure

We next sought to determine the structure of a peptide mimetic complex and utilised the commercially available inhibitors PCK (Pro-Phe-Arg) and PPACK (Phe-Pro-Arg). These did not co-crystallize with βFXIIa^{His} so we engineered the MBP fusion construct to facilitate structure determination with inhibitors. MBP-βFXIIa^{His} crystals were obtained through co-crystallization only with the PPACK inhibitor and the structure was determined by molecular replacement. Two copies of the MBP-βFXIIa^{His} fusion were placed in the resulting high-quality electron density map and the refined model (Table 2) showed PPACK bound to the active site, and the FXII Asn74 N-linked glycan residues are also modelled (Fig. S3). Fig. 4*A* shows the overall structure of the MBP-βFXIIa^{His} fusion with the βFXIIa^{His} protease domain (coloured as a rainbow) positioned on top of the MBP (colored in green). The amino acid linker of sequence AlaAlaAlaAlaSer connecting the C-terminus of MBP to the N-terminus of βFXIIa^{His} is partially observed in the electron density (Fig. 4*A*, S3) and is assumed to be

flexible (shown as a black dotted line). The 5 amino acids from the HCR fragment including the Cys340-Cys467 disulfide are also observed (Fig. 4*A*).

Unexpectedly multiple interactions are formed between MBP and βFXIIa^{His} extending over a surface area of 654.8 Å^2, (calculated using PISA) (Fig. 4*B*). For both copies of MBP-βFXIIa^{His} the same organisation of MBP and βFXIIa^{His} is observed and is superposable with an r.m.s.d. of 0.688 Å (for 614 C-α atoms). This MBP-βFXIIa^{His} interface occurs between three irregular loops from the FXII protease domain β-barrels and helices α2, α12 of MBP. From MBP helix α2 Lys46 forms a salt bridge to Asp111 from FXII, MBP Lys42 forms hydrogen bonds to the main chain carbonyl groups of residues Glu78 and Cys80 and MBP Gln49 (helix α2) forms a hydrogen bond to the carbonyl of FXII Leu114 (Fig. 4*B*). A second series of interactions originates from MBP helix α12 with a salt bridge between MBP Arg354 and FXII Glu71 and hydrogen bonding interactions formed between the side chain groups of MBP Asn349 and Thr345 with the main chain nitrogen of FXII Arg24 and the side chain of Tyr117, respectively. The MBP Tyr341 side chain forms a hydrophobic interaction with FXII Pro116, MBP Asp207 forms a salt bridge to FXII Arg72 and the MBP Asn150 side chain hydrogen bonds to the FXII Glu78 side chain. These interactions result in the FXII protease covering the entrance to the MBP pocket where maltose binds and may explain why this additive was required for the purification and crystallization experiments.

In the MBP-βFXIIa^{His}-PPACK structure, the Arg-Pro residues form characteristic interactions with the FXIIa S1 and S2 pockets as observed in the thrombin complex crystal structure, however, the P3 PPACK Phe residue projects away from the protease and is involved in a crystal contact. This differs from the thrombin-PPACK complex where the Phe residue is inserted in the S3 pocket (Fig. 4*C*) (Bode *et al.*, 1989).

3.6 The complex of βFXIIa^{His} with inhibitor EcTI

We next chose to characterise whether the recombinant insect cell expressed βFXIIa^{His} would interact with a kunitz type protease inhibitor and perform a comparison with the plasma purified βFXIIa, which will have a larger glycan structure. EcTI is a trypsin inhibitor derived from the plant *Enterolobium contortisiliquum,* which has been reported as inhibiting FXIIa activity and a crystal structure of EcTI in complex with trypsin is available (Zhou *et al.*, 2013). To characterise the βFXIIa–EcTI complex gel filtration (Superdex 200 Increase) chromatography was performed with an equimolar mixture of βFXIIa^{His} and EcTI, with the elution profile shown in Fig. 5*A*. This revealed a βFXIIa^{His}-EcTI complex of approximately 43 kDa with an excess of EcTI identified as a separate peak (~20 kDa) confirmed by SDS-PAGE of the resulting fractions (Fig. 5*A*).

Next, surface plasmon resonance (SPR) was performed to quantify the βFXIIa^{His}-EcTI binding. As shown in Fig. 5*B*, SPR binding curves demonstrated concentration-dependent binding. The data were fitted using the 1:1 binding model with mass transfer to derive values of K_D, k_a and k_d. The K_D, k_a and k_d for EcTI binding to the plasma purified βFXII^{His} were 2.68×10^{-7} M, 1.02×10^{4} M^{-1} s^{-1}, and 2.71×10^{-3} s^{-1}, respectively. These values are comparable to EcTI binding to recombinant βFXII^{His} and the K_D, k_a and k_d values were 3.38×10^{-7} M, 8.17×10^{3}, M^{-1} s^{-1}, and 2.76×10^{-3} s^{-1}, respectively.

To understand the molecular basis of the Kunitz-type inhibitor complex with βFXIIa we performed molecular docking with the crystal structures of βFXIIa^{His} and EcTI (pdb code: 4J2Y) using the program Cluspro (Kozakov *et al.*, 2017). Inspection of the top ranked docking solution identified a canonical complex whereby the EcTI residue Arg64 was inserted into the S1 pocket of βFXIIa^{His} with scores of -769.4 and -962.1, representing centre and lowest energy scores, respectively, for 274 members in the largest cluster size (the second ranked solution had 101 members), and no steric clashes were observed.

Fig. 6*A,B* shows a cartoon diagram of the predicted βFXIIa^{His}-EcTI complex and Fig. 6*C* a surface charge distribution of βFXIIa^{His} with EcTI as a cartoon diagram. Table S1 lists the key electrostatic and hydrogen bond interactions at the interface, burying a total surface area of 1207 Å^2 (pdbe-PISA) in comparison to 988 Å^2 observed in the trypsin-EcTI complex (pdb code: 4J2K). The EcTI Arg64 side-chain extends into the S1 pocket of the FXIIa catalytic domain forming a salt bridge with the S1 pocket residue Asp189 side-chain carboxyl (chymotrypsin numbering used) (Fig. 6*B*). Additional interactions flanking the P1 Arg64 include Pro63 interacting with the S2 pocket (Fig. 6*D*) which is similar to interactions observed in the MBP-βFXIIa^{His}-PPACK complex.

3.7 P2 substrate preference of βFXIIa

FXIIa cleaves three substrates FXII, PK, FXI and substrate selection is driven by a combination of exosite and subsite interactions that are poorly understood. An alignment of the sequences in the area of the cleavage site is shown in Table 3. All have Arg at the P1 position, but at the P2 position there is greater variability as FXI has Pro and FXII, PK both have a Thr residue. High throughput protease specificity profiling using a diverse short peptide based fluorogenic substrate microarray confirmed the requirement of Arg at P1 and revealed a FXIIa preference for Phe and Thr at the P2 position, which is consistent with the sequence Pro-Phe-Arg of the widely used FXIIa substrate S-2302 (Gosalia *et al.*, 2006, Zhou *et al.*, 2013). Table 3 shows FXII inhibitors follow this preference for Phe and Thr at the P2 position (PCK, infestin-4, and bicyclic peptide (Middendorp *et al.*, 2017)) and also mimic its substrate FXI with Pro in the P2 (CTI and EcTI). This more open S2 pocket observed in both MBP-βFXIIa^{His} and βFXIIa^{His} presents a framework that explains the FXIIa substrate preference for a bulky Phe sidechain at the P2 position. Why the natural substrates utilise Thr at P2 more commonly and do not utilise Phe at P2 is unclear, but as discussed below, it is likely that exogenous inhibitor infestin-4 does utilise the open S2 pocket to interact with a P2 Phe sidechain.

3.7 Docking of βFXIIa^{His} with Infestin-4

The exogenous inhibitor Infestin-4 is derived from the blood sucking kissing bug and has been characterised in detail as a FXIIa inhibitor which has also been optimised for selectivity. Previous studies have performed docking on Infestin-4 using a model of FXIIa and presented a complex which places the Infestin-4 P2 Phe9 sidechain in the S1' pocket and the P1' Asn11 in the S2 pocket (Kolyadko *et al.*, 2015). A crystal structure of Infestin-4 was utilised in docking simulations with the βFXIIa^{His} crystal structure. Figure 7 shows the highest scoring complex with the Infestin-4 P1 Arg and P2 Phe inserted into the βFXIIa^{His} S1 and S2 pocket respectively. Interestingly three highly complementary salt bridges are formed between infestin-4 residues Lys36, Lys21 and Arg40 with βFXIIa^{His} residues Asp222 from 220-loop and E151B, E151E from the 140-loop, respectively. This orientation is the reverse of the Kolyadko *et al.,* 2015 Infestin-4:FXIIa docked complex and highlights the importance of consdering the βFXIIa^{His} crystal structure with an open S2 pocket as a template for docking studies.

4. Discussion

Efforts are currently underway to design selective inhibitors for FXIIa as novel treatments for thrombotic and inflammatory disorders (Middendorp *et al.*, 2017, Kolyadko *et al.*, 2015, Krupka *et al.*, 2016, Campos *et al.*, 2012). To investigate the FXIIa protease structure in complex with peptidomimetic inhibitors we developed vectors and an insect cell based protocol for recombinant expression and purification of the βFXIIa protease domain achieving yields suitable for enzyme kinetic and protein crystallography experiments. MBP-βFXIIa^{His} crystals were obtained through co-crystallization experiments with the PPACK inhibitor. PPACK is a selective inhibitor of thrombin and a comparison of the PPACK-MBP-βFXIIa^{His} structure with the reported PPACK-thrombin complex (Bode *et al.*, 1989) revealed

the PPACK P3 Phe residue forms interactions with residues in the S3 pocket in the thrombin complex, whereas in the MBP-βFXIIa^{His} complex the Phe residue is projecting outwards and not forming significant interactions with contacts exclusively through the PPACK Arg and Pro residues which is reminiscent of the FIXa-PPACK complex (Kristensen *et al.*, 2016).

A second construct βFXIIa^{His} crystallized in the absence of an inhibitor, but here the active conformation was stabilized by Thr-Arg residues from an adjacent molecule in the crystal. This Thr-Arg sequence mimics the natural FXIIa substrate P1-P2 cleavage site residues present in prekallikrein and FXII activation loop sequences. Previous studies utilizing tripeptide substrates have determined the FXIIa optimal residues at P2 to be Phe or Thr, and the substrate S-2302 widely used to monitor FXIIa activity has the sequence Pro-Phe-Arg. The MBP-βFXIIa^{His}-PPACK (Pro-Arg) and βFXIIa^{His}-Thr-Arg substrate mimetic structures reveal that the Tyr99 residue alters its position from packing against His57 (observed in the zymogen FXIIc structure), which effectively opens up the S2 pocket. This movement of Tyr99 comes at the expense of the S3 pocket, which is partially occluded by packing of Tyr99 against Trp215.

Infestin-4 and a bicyclic peptide are selective protein/peptide based inhibitors of FXIIa which both utilise Phe residue at the P2 position (Middendorp *et al.*, 2017, Kolyadko *et al.*, 2015, Krupka *et al.*, 2016, Campos *et al.*, 2012). The more open S2 pocket provides a model for the preference of the bulky Phe residue at P2 and future studies involving a crystal structure with bound substrate mimetic or exogenous inhibitor will be required to demonstrate how this pocket is utilised. A series of crystal structures have been reported for protease domains that have a similar S2 pocket and residue Tyr99 (FXa, tPa and FIXa), however, most of these have the closed configuration of S2 with Tyr99 packing against His57, implying this transiently opens only in the presence of a substrate mimetic. A comparison of the recombinant βFXIIa^{His}-Thr-Arg structure with the benzamidine-βFXIIa structure reveals the latter also has a closed S2 pocket (Dementiev *et al.*, 2018).

The role of the distinctive negatively charged ridges in the 140 and 220-loops observed surrounding the FXII active site is unknown(Pathak *et al.*, 2015), but the substrates FXI (Mohammed *et al.*, 2018), PK (Hooley *et al.*, 2007) and the natural inhibitor serpin C1 inhibitor have positively charged residues (Huntington, 2011, Beinrohr *et al.*, 2007, Hamad *et al.*, 2017) that may provide electrostatic interactions with βFXIIa. Overall, these βFXIIa structures provide a framework to understand substrate selectivity and for rational design of inhibitors to target thrombotic and inflammatory diseases.

Author contributions

R Manna determined the MBP-βFXIIa^{His} structure and M Pathak determined the βFXIIa^{His} structure and contributed to preparation of the draft manuscript. D Belviso and C Li performed crystallographic refinement analysis. R Bonturi and Maria Luiza V Oliva isolated and purified EcTI. B Hamad performed enzyme kinetic data analysis. J Emsley, LV Dekker, PM Fischer, I Dreveny conceived and designed the study, generated, analyzed and collected/interpreted data, wrote and revised the intellectual content and generated the final manuscript version.

Acknowledgements

Crystallographic data were obtained with the support of Diamond Light Source using the beamlines I04 and I24.

PDB reference The atomic coordinates and structure factors of βFXIIa^{His}, (pdb code: 6GT6) and MBP-βFXIIa^{His} (pdb code: 6QF7) have been deposited in the Protein Data Bank (http://www.rcsb.org).

Funding information

This work was supported by the British Heart Foundation studentship award FS/11/61/28941 to Rosa Manna and program grant programme Grant no. RG/12/9/29775 to JE. Also grant Fundação de Amparo à Pesquisa do Estado de São (FAPESP) [2017/07972-9 and 2017/06630-7]; Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) [23038.007762/2014-32, AUXPE 140/2015] to MLVO. The authors declare that they have no conflicts of interest with the contents of this article.

Abbreviations. FXII, Factor XII, PK, Prekallikrein; PDB, Protein Data Bank, CTI, Corn Trypsin inhibitor, MBP, Maltose binding protein

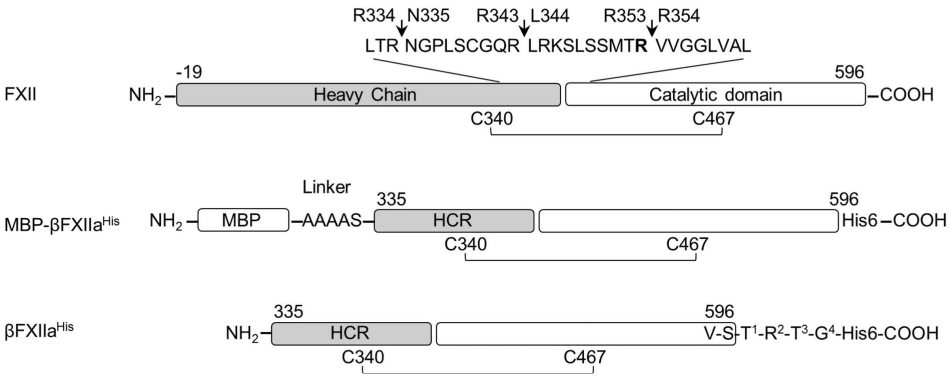

Figure 1. FXII domain organization and protein expression construct designs. Schematic representation of the FXII domain organization (top) with the disulfide bridge (Cys340-Cys467) between heavy chain and catalytic domain is shown. The principal activating cleavage site is depicted as an arrow between residues Arg353-Val354. Other cleavage sites are depicted with Arg334-Asn335 resulting in the formation of βFXIIa. Below this is the MBP-βFXIIa^{His} and βFXIIa^{His} constructs used to generate recombinant βFXIIa. HCR denotes the 9 amino acid heavy chain remnant found in βFXIIa.

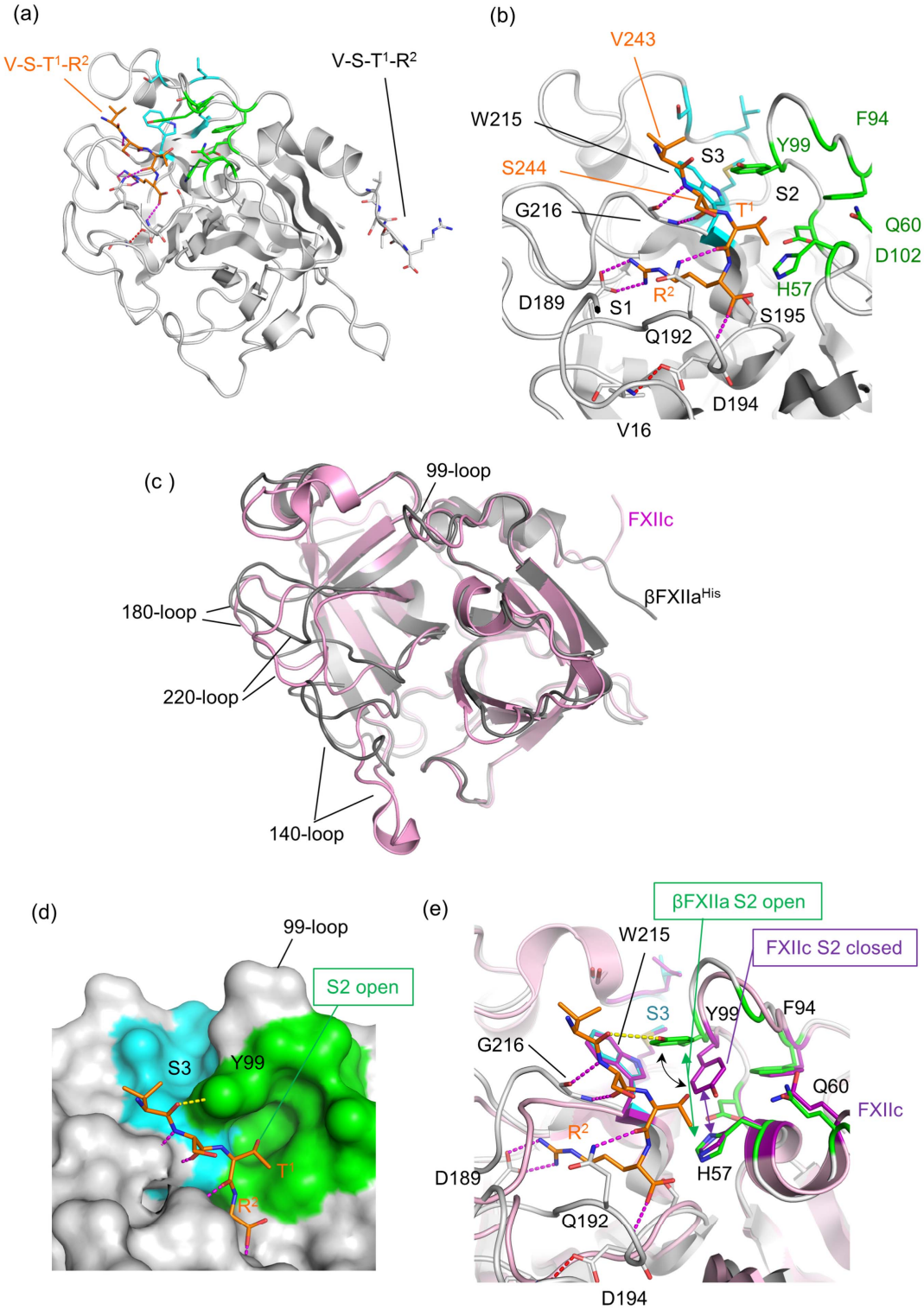
(a)
V-S-T¹-R²
V-S-T¹-R²
(b)
V243
W215
F94
Y99
S3
S244
S2
G216
T¹
Q60
D102
H57
R²
D189
S1
S195
Q192
D194
V16
(c)
99-loop
FXIIc
βFXIIa^His
180-loop
220-loop
140-loop
(d)
99-loop
S2 open
S3
Y99
T¹
R²
(e)
βFXIIa S2 open
W215
FXIIc S2 closed
S3
Y99
F94
G216
Q60
FXIIc
R²
D189
H57
Q192
D194

Figure 2. Structure of βFXIIa^{His}. (a) Cartoon diagram of the βFXIIa^{His} active site is shown occupied by four amino acids (sticks colored orange) from the C-terminus of an adjacent molecule in the crystal. Thr^{1}-Arg^{2} are numbered thus as they are not derived from the FXII sequence but from the C-terminal regions of the expression construct. Residues from the S2 pocket are shown as sticks in green and S3 in cyan. (b) A close up of the structure in *A* shown with electrostatic interactions between V-S-T-R residues and residues from the βFXIIa^{His} S1, S2 and S3 pockets illustrated as purple dotted lines. (c) Cartoon diagram of βFXIIa^{His} (grey) superposed onto the FXIIc structure (purple) with key loops altering conformation labeled (140-loop represents residues in the range of residue 140 using chymotrypsin numbering which is also known as the autolysis loop). (d) Surface representation of the S2 and S3 pockets in a similar view as shown in *B.* Close contact of the Tyr99 hydroxyl with the V-S-T-R main chain carbonyl is shown as a yellow dotted line. (e) The S2 pocket in βFXIIa^{His} with V-S-T-R residues (orange sticks) is shown superposed onto the FXIIc structure (purple) illustrating the positional change of residue Tyr99 annotated as an open or closed S2 pocket.

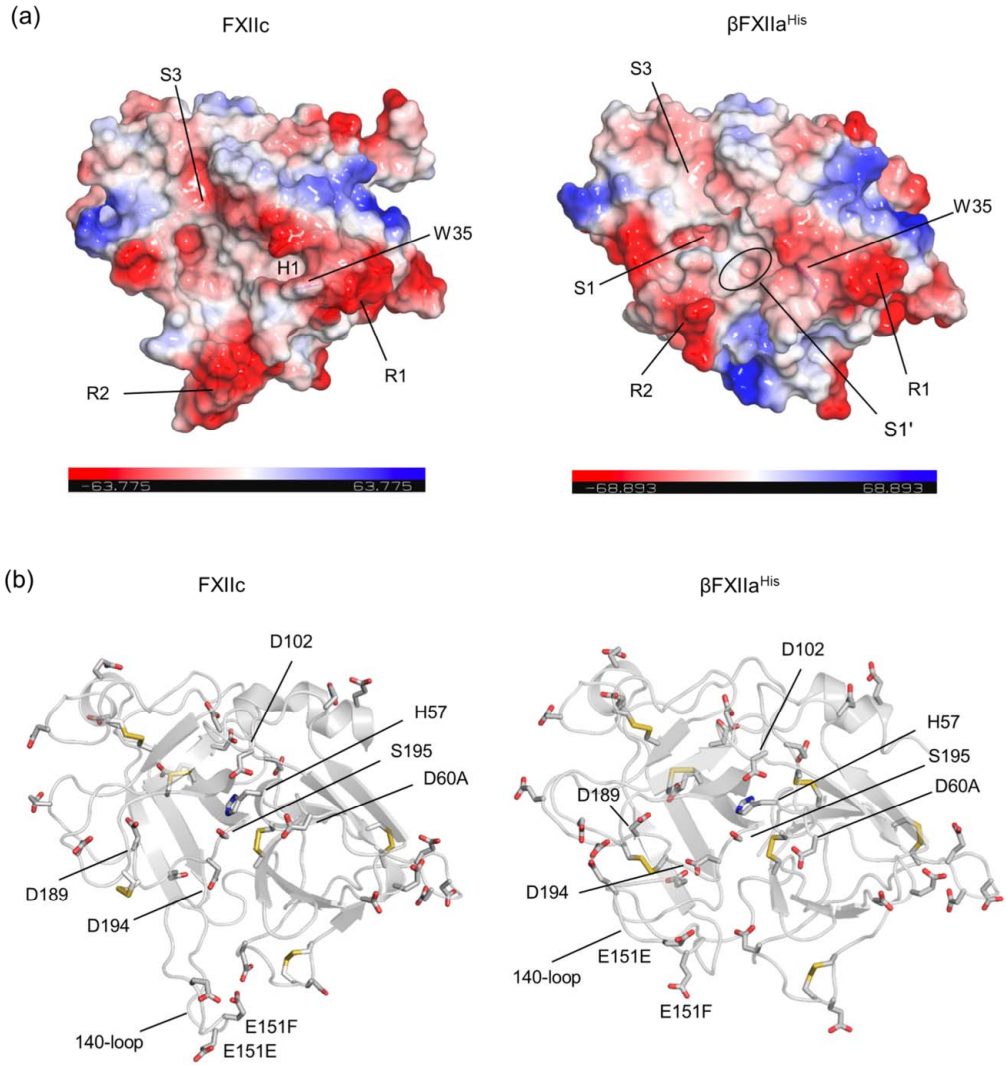

Figure 3. Changes in the overall surface charge distribution in the βFXIIa^{His} structure compared to the zymogen like FXIIc structure. (a) Electrostatic (red=negative; blue= positive) charge distribution in both βFXIIa (right) and FXIIc (left) with a large movement of the ridge of negative charge labeled R2 is shown. (b) Cartoon diagram showing the distribution of acidic residues highlighted as sticks in both the zymogen FXIIc (left) and the active βFXIIa (right).

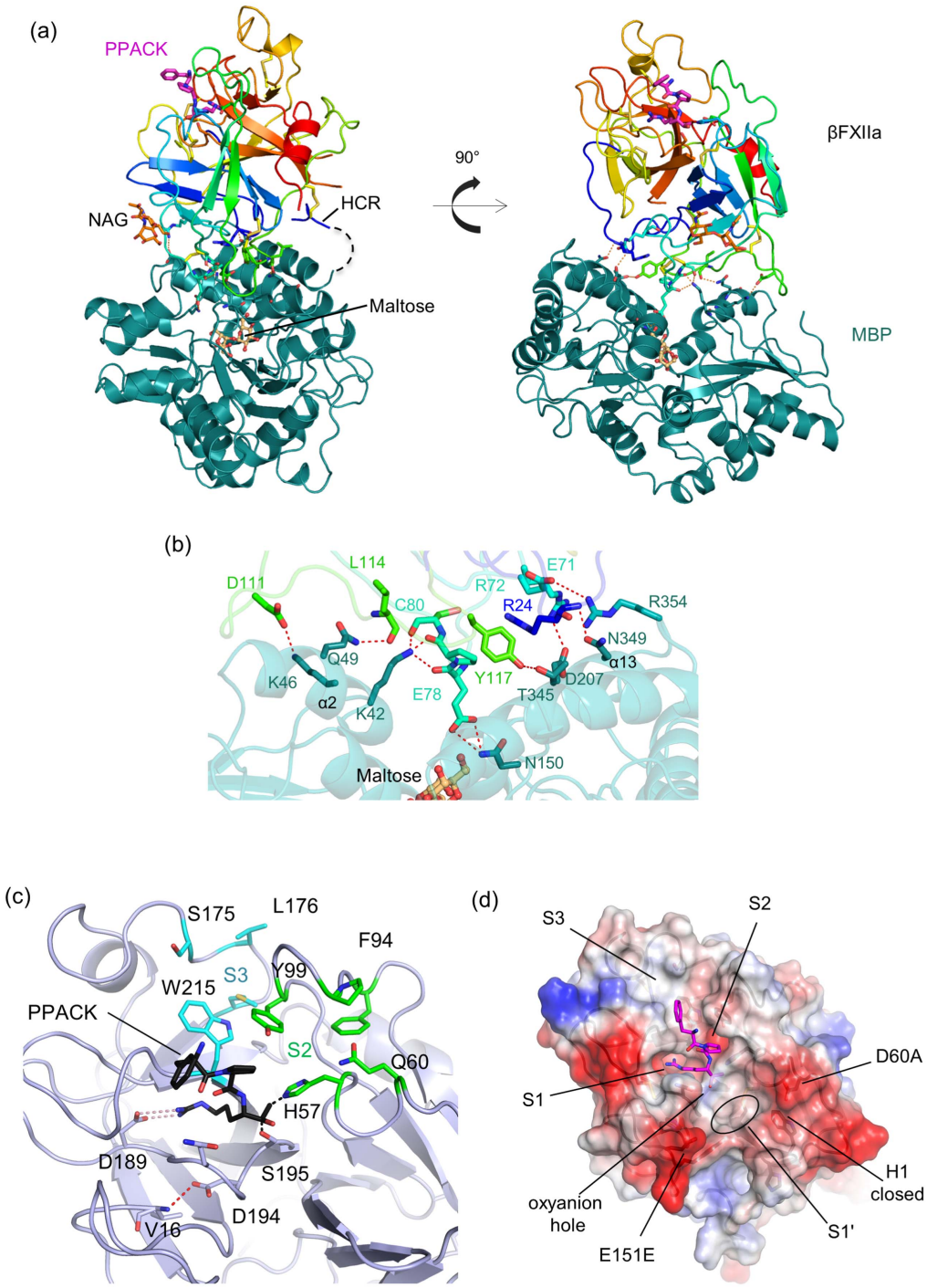
(a)
PPACK
NAG
HCR
Maltose
90°
βFXIIa
MBP
(b)
D111
L114
C80
R72
E71
R24
R354
N349
α13
Q49
K46
α2
K42
E78
Y117
D207
T345
N150
Maltose
(c)
S175
L176
F94
Y99
S3
W215
PPACK
S2
Q60
H57
D189
S195
D194
V16
(d)
S3
S2
D60A
S1
H1
closed
oxyanion
hole
E151E
S1'

Figure 4. Structure of the MBP-βFXIIa^{His} crystal structure complexed with PPACK. (a) Cartoon representation of the crystal structure of the MBP-βFXIIa^{His}-PPACK complex. MBP shown in dark green and βFXIIa^{His} in rainbow with HCR indicated and N-linked glycan residues as sticks in orange and PPACK (purple). The maltose bound to MBP is drawn in ball and stick in yellow. (b) Ribbon diagram of the interactions occurring at the interface between MBP and the βFXIIa^{His} fusion partner. The interacting residues are drawn in sticks with electrostatic interactions depicted as dashed lines. (c) Detailed interactions of PPACK (black) with key residues shown as sticks forming interactions with the S1 pocket (dashed lines). The dotted circle shows the oxyanion hole and Asp189 residue forming the base of the S1 pocket is also shown. (d) MBP-βFXIIa^{His} protease surface charged representation (negative=red, positive=blue). The PPACK P1 arginine is bound in the S1 pocket (purple sticks).

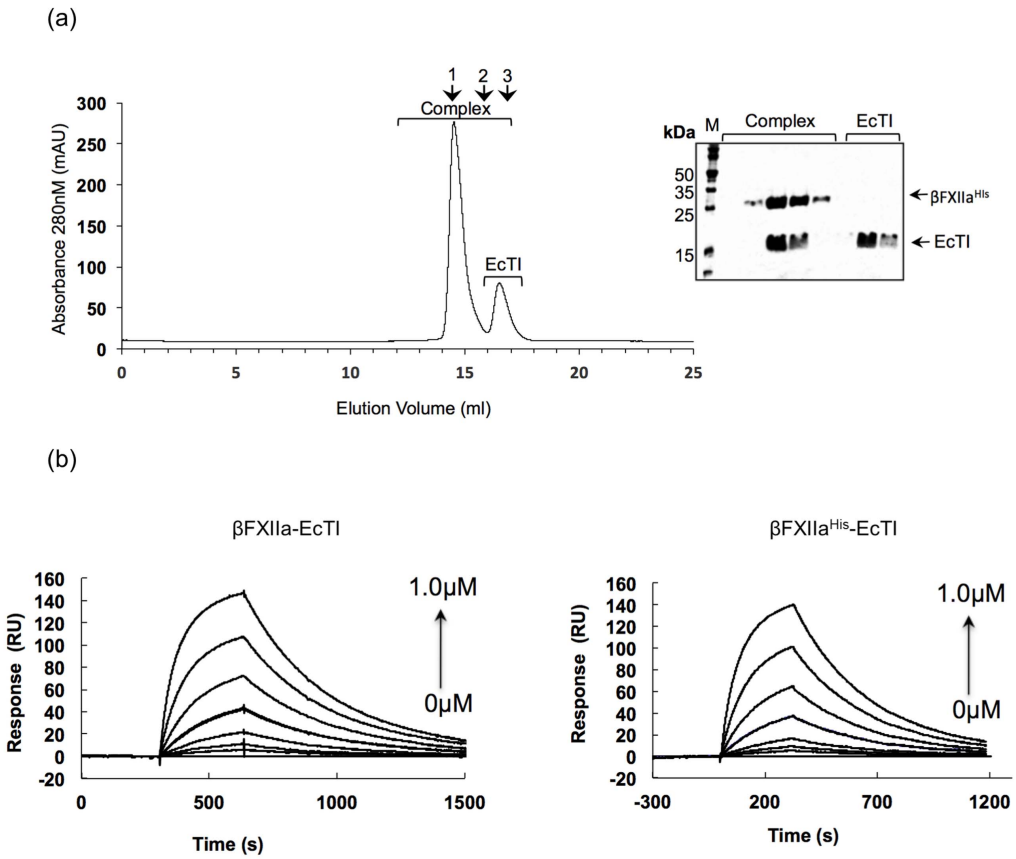

Figure 5. βFXIIa binding to kunitz type inhibitor EcTI. (a) The βFXIIa^{His}-EcTI complex was analyzed by gel filtration. Arrows numbered 1, 2, and 3 indicate the elution volume of molecular weight standards of 43 kDa, 22 kDa and 13.7 kDa respectively. The panel on the left shows the coomassie stained SDS-PAGE gel under reducing conditions with different lanes representing eluted fractions from gel filtration of the βFXIIa^{His}-EcTI complex. (b) SPR analysis showing the commercial βFXIIa (left) and recombinant βFXIIa^{His} (right) binding to EcTI. SPR sensorgrams illustrate the βFXIIa interactions with immobilized EcTI at various βFXIIa concentrations. The arrow indicates increasing βFXIIa concentrations used at 0, 15.625, 31.25, 62.5, 125, 250, 500, 1000 nM with response units (RU) on the y axis.

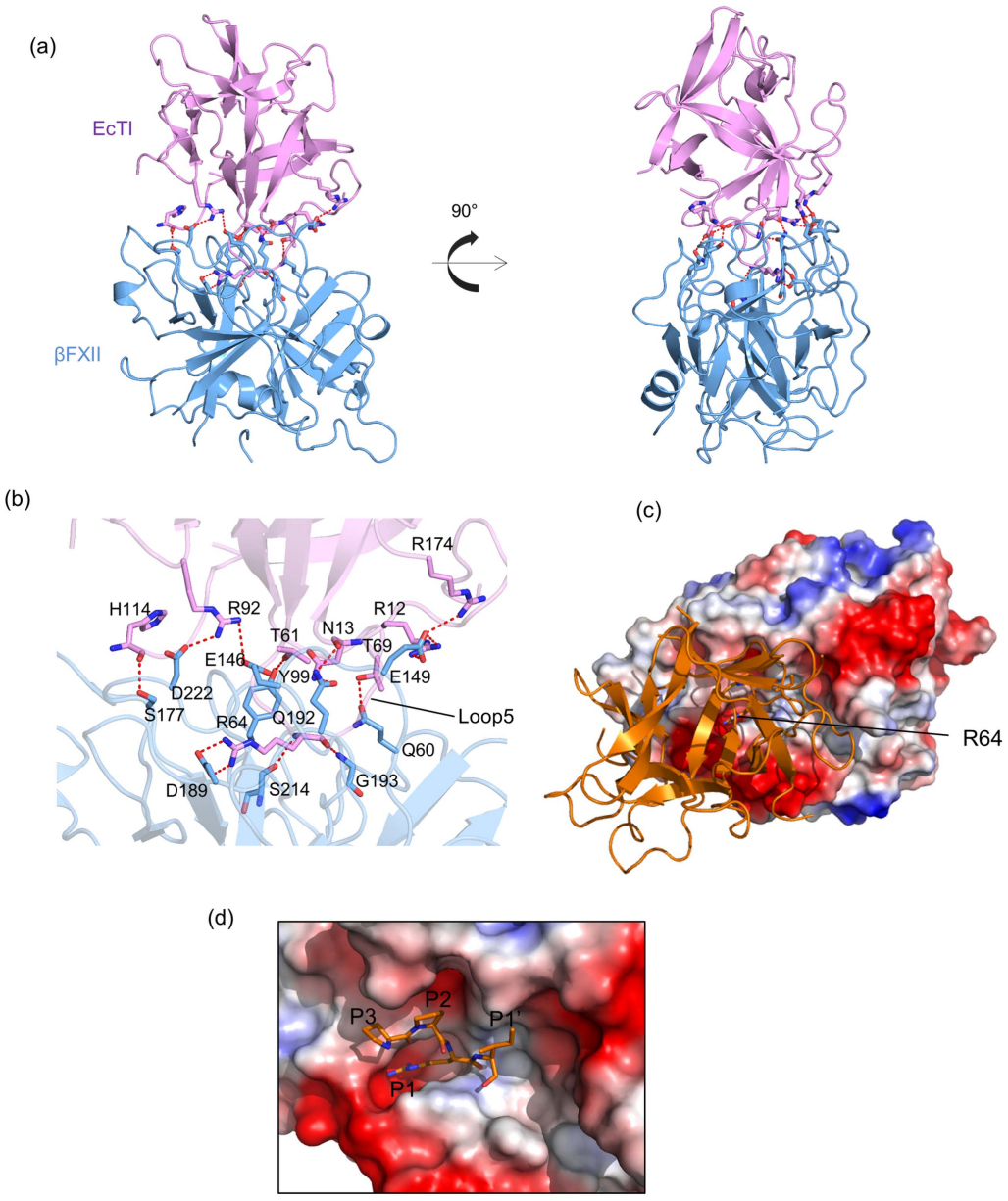

Figure 6. Molecular docking of the inhibitor EcTI with βFXIIa^{His}. (a) Cartoon representation of the Cluspro docked complex of βFXIIa^{His} (blue) with EcTI (pink) shown as two rotated views with interfacial residues as sticks. (b) Close up view of the interface between EcTI and βFXIIa^{His} with the interacting residues drawn in sticks and electrostatic interactions as dashed lines. (c) Charged surface representation of βFXIIa^{His} with a cartoon representation of the docked EcTI structure (orange). (d) Zoomed in view of the charged surface representation of the active site of βFXIIa^{His} with bound EcTI reactive loop P3-P1' residues shown as sticks (orange).

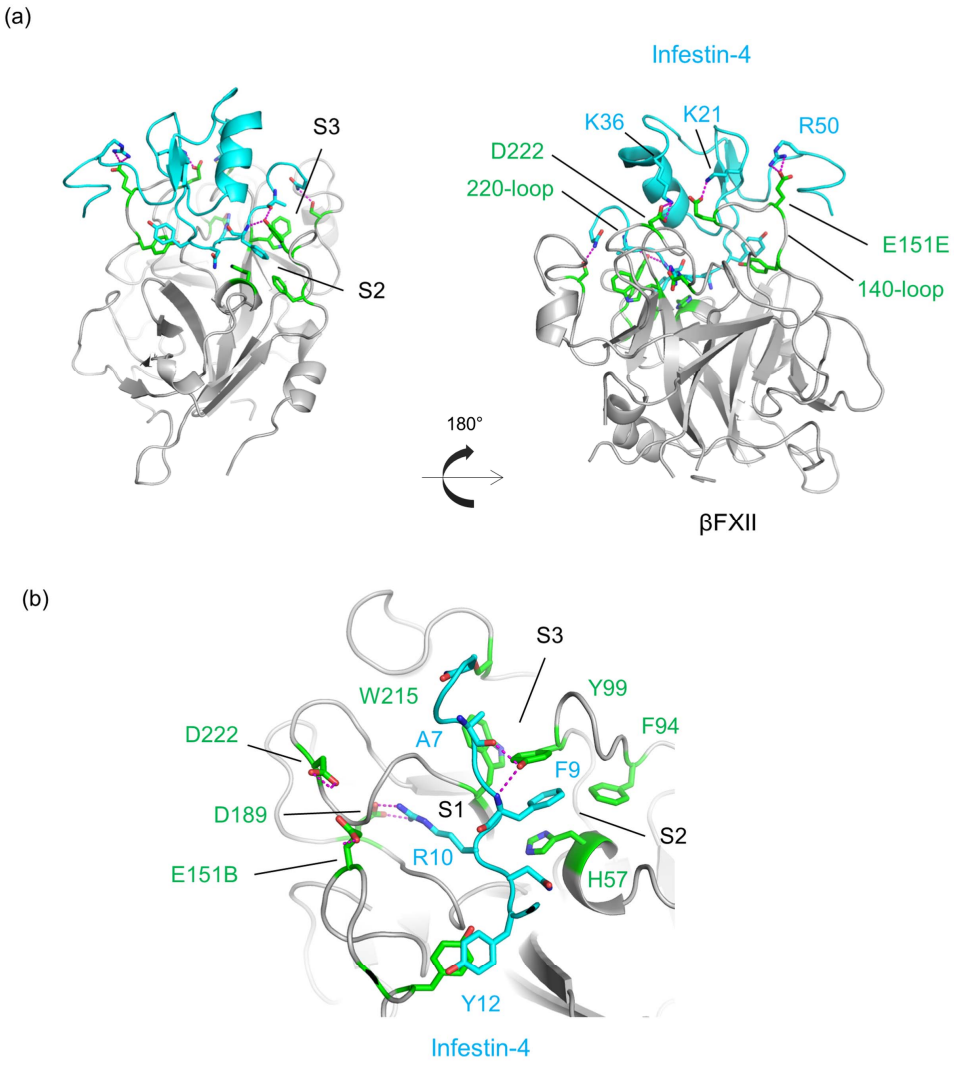

Figure 7. Molecular docking of the Infestin-4 crystal structure (pdb code:2ERW) with βFXIIa^{His}. (a) Cartoon representation of the Cluspro docked complex of βFXIIa^{His} (grey) with Infestin-4 (cyan) shown as two rotated views with interfacial residues as sticks. (b) Close up view of the interface between Infestin-4 and βFXIIa^{His} with the interacting residues drawn in sticks and electrostatic interactions as dashed lines.

Table 1. Enzyme kinetic parameters for FXIIa samples cleaving substrate S-2302

Sample	V_{max} (pmol/s)	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m (L mol^{-1} s^{-1})
MBP-βFXIIa^{His}	24.40±0.60	17.43±0.83	0.10±0.02	174300
βFXIIa^{His}	32.93±2.70	21.75±1.86	0.20±0.04	108750
βFXIIa	34.20±3.24	24.44±0.75	0.16±0.01	101880
αFXIIa	21.72±1.04	15.52±0.74	0.17±0.02	91294

Values are reported with standard error of the regression analysis calculated with GRAPHPAD PRISM using a non-linear regression with the Michaelis-Menten equation.

Table 2. Data collection and refinement statistics

	MBP-βFXIIa^{His} PPACK	βFXIIa^{His}
Data Collection		
Space group	P 4_1 2_1 2	P 4_1 2_1 2
Unit cell dimensions		
a, b, c (Å)	131.5, 131.5, 238.4	106.9, 106.9, 66.0
α, β, γ (°)	90, 90, 90	90, 90, 90
Wavelength (Å)	0.9795	0.97951
Resolution (Å)	4.0	2.54
R_{merge}	0.189 (0.468)	0.218 (0.821)
I/σI	3.8 (1.7)	5.6 (1.7)
Completeness (%)	94.8 (95.7)	99.9 (100)
Multiplicity	2.5	6.3
CC ½	0.95 (0.61)	0.985 (0.35)
Unique reflections	17015 (4810)	13131 (938)
Refinement		
R_{work}	0.297 (0.368)	0.213(0.219)
R_{free}	0.356 (0.401)	0.265 (0.269)
B-factor (A^2)	38.3	42.0
Stereo chemical r.m.s.d.		
Bond length (Å)	0.011	0.010
Bond angle (°)	1.487	1.439
Ramachandran plot		
Most favored (%)	91.2	95.26
Allowed (%)	8.8	4.74
Outliers (%)	0.0	0.0

r.m.s.d., root mean square deviation. Values in parenthesis correspond to the highest-resolution shell.

Table 3. Amino acid sequences of FXIIa substrate and inhibitor reactive loops.

	P3	P2	**P1**	P1′	P2′
Substrates					
PK	S	T	**R**	I	V
FXII	M	T	**R**	V	V
FXI	K	P	**R**	I	V
Inhibitors					
PCK	P	F	**R**		
Infestin-4	C	F	**R**	N	Y
Infestin-4mB	C	T	**R**	N	F
CTI	G	P	**R**	L	P
Bicyclic-pep	C	F	**R**	L	P
EcTI	P	P	**R**	I	A
PPACK	F	P	**R**		

References:

Baglin, T. P., Carrell, R. W., Church, F. C., Esmon, C. T. & Huntington, J. A. (2002). *Proc Natl Acad Sci U S A* **99**, 11079-11084.

Batista, I. F., Oliva, M. L., Araujo, M. S., Sampaio, M. U., Richardson, M., Fritz, H. & Sampaio, C. A. (1996). *Phytochemistry* **41**, 1017-1022.

Beinrohr, L., Harmat, V., Dobo, J., Lorincz, Z., Gal, P. & Zavodszky, P. (2007). *J Biol Chem* **282**, 21100-21109.

Bender, L., Weidmann, H., Rose-John, S., Renne, T. & Long, A. T. (2017). *Front Immunol* **8**, 1115.

Bode, W., Mayr, I., Baumann, U., Huber, R., Stone, S. R. & Hofsteenge, J. (1989). *Embo j* **8**, 3467-3475.

Bokhove, M., Sadat Al Hosseini, H., Saito, T., Dioguardi, E., Gegenschatz-Schmid, K., Nishimura, K., Raj, I., de Sanctis, D., Han, L. & Jovine, L. (2016). *J Struct Biol* **194**, 1-7.

Campos, I. T., Souza, T. A., Torquato, R. J., De Marco, R., Tanaka-Azevedo, A. M., Tanaka, A. S. & Barbosa, J. A. (2012). *Acta Crystallogr D Biol Crystallogr* **68**, 695-702.

Campos, I. T., Tanaka-Azevedo, A. M. & Tanaka, A. S. (2004). *FEBS Lett* **577**, 512-516.

Cichon, S., Martin, L., Hennies, H. C., Muller, F., Van Driessche, K., Karpushova, A., Stevens, W., Colombo, R., Renne, T., Drouet, C., Bork, K. & Nothen, M. M. (2006). *Am J Hum Genet* **79**, 1098-1104.

Dementiev, A., Silva, A., Yee, C., Li, Z., Flavin, M. T., Sham, H. & Partridge, J. R. (2018). *Blood Adv* **2**, 549-558.

Emsley, P. & Cowtan, K. (2004). *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132.

Evans, P. (2006). *Acta Crystallogr D Biol Crystallogr* **62**, 72-82.

Evans, P. R. & Murshudov, G. N. (2013). *Acta Crystallogr D Biol Crystallogr* **69**, 1204-1214.

Gong, L., Liu, M., Zeng, T., Shi, X., Yuan, C., Andreasen, P. A. & Huang, M. (2015). *J Biol Chem* **290**, 25795-25804.

Gosalia, D. N., Denney, W. S., Salisbury, C. M., Ellman, J. A. & Diamond, S. L. (2006). *Biotechnol Bioeng* **94**, 1099-1110.

Hamad, B. K., Pathak, M., Manna, R., Fischer, P. M., Emsley, J. & Dekker, L. V. (2017). *J Thromb Haemost* **15**, 1818-1828.

Hess, R., Wujak, L., Hesse, C., Sewald, K., Jonigk, D., Warnecke, G., Fieguth, H. G., de Maat, S., Maas, C., Bonella, F., Preissner, K. T., Weiss, B., Schaefer, L., Kuebler, W. M., Markart, P. & Wygrecka, M. (2017). *Thromb Haemost* **117**, 1896-1907.

Hooley, E., McEwan, P. A. & Emsley, J. (2007). *J Thromb Haemost* **5**, 2461-2466.

Huntington, J. A. (2011). *J Thromb Haemost* **9 Suppl 1**, 26-34.

Ivanov, I., Matafonov, A., Sun, M. F., Cheng, Q., Dickeson, S. K., Verhamme, I. M., Emsley, J. & Gailani, D. (2017). *Blood* **129**, 1527-1537.

Iwaki, T. & Castellino, F. J. (2008). *Cytotechnology* **57**, 45-49.

Jin, T., Chuenchor, W., Jiang, J., Cheng, J., Li, Y., Fang, K., Huang, M., Smith, P. & Xiao, T. S. (2017). *Sci Rep* **7**, 40991.

Kabsch, W. (2010). *Acta Crystallogr D Biol Crystallogr* **66**, 125-132.

Kleinschnitz, C., Stoll, G., Bendszus, M., Schuh, K., Pauer, H. U., Burfeind, P., Renne, C., Gailani, D., Nieswandt, B. & Renne, T. (2006). *J Exp Med* **203**, 513-518.

Kolyadko, V. N., Lushchekina, S. V., Vuimo, T. A., Surov, S. S., Ovsepyan, R. A., Korneeva, V. A., Vorobiev, II, Orlova, N. A., Minakhin, L., Kuznedelov, K., Severinov, K. V., Ataullakhanov, F. I. & Panteleev, M. A. (2015). *PLoS One* **10**, e0144940.

Kozakov, D., Hall, D. R., Xia, B., Porter, K. A., Padhorny, D., Yueh, C., Beglov, D. & Vajda, S. (2017). *Nat Protoc* **12**, 255-278.

Kristensen, L. H., Olsen, O. H., Blouse, G. E. & Brandstetter, H. (2016). *Biochem J* **473**, 2395-2411.

Krupka, J., May, F., Weimer, T., Pragst, I., Kleinschnitz, C., Stoll, G., Panousis, C., Dickneite, G. & Nolte, M. W. (2016). *PLoS One* **11**, e0146783.

Larsson, M., Rayzman, V., Nolte, M. W., Nickel, K. F., Bjorkqvist, J., Jamsa, A., Hardy, M. P., Fries, M., Schmidbauer, S., Hedenqvist, P., Broome, M., Pragst, I., Dickneite, G., Wilson, M. J., Nash, A. D., Panousis, C. & Renne, T. (2014). *Sci Transl Med* **6**, 222ra217.

Maas, C. & Renne, T. (2018). *Blood*.

MacQuarrie, J. L., Stafford, A. R., Yau, J. W., Leslie, B. A., Vu, T. T., Fredenburgh, J. C. & Weitz, J. I. (2011). *Blood* **117**, 4134-4141.

Mahoney, W. C., Hermodson, M. A., Jones, B., Powers, D. D., Corfman, R. S. & Reeck, G. R. (1984). *J Biol Chem* **259**, 8412-8416.

Matafonov, A., Leung, P. Y., Gailani, A. E., Grach, S. L., Puy, C., Cheng, Q., Sun, M. F., McCarty, O. J., Tucker, E. I., Kataoka, H., Renne, T., Morrissey, J. H., Gruber, A. & Gailani, D. (2014). *Blood* **123**, 1739-1746.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J Appl Crystallogr* **40**, 658-674.

Middendorp, S. J., Wilbs, J., Quarroz, C., Calzavarini, S., Angelillo-Scherrer, A. & Heinis, C. (2017). *J Med Chem* **60**, 1151-1158.

Mohammed, B. M., Matafonov, A., Ivanov, I., Sun, M. F., Cheng, Q., Dickeson, S. K., Li, C., Sun, D., Verhamme, I. M., Emsley, J. & Gailani, D. (2018). *Thromb Res* **161**, 94-105.

Moon, A. F., Mueller, G. A., Zhong, X. & Pedersen, L. C. (2010). *Protein Sci* **19**, 901-913.

Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Crystallogr D Biol Crystallogr* **67**, 355-367.

Nickel, K. F., Long, A. T., Fuchs, T. A., Butler, L. M. & Renne, T. (2017). *Arterioscler Thromb Vasc Biol* **37**, 13-20.

Pathak, M., Wilmann, P., Awford, J., Li, C., Hamad, B. K., Fischer, P. M., Dreveny, I., Dekker, L. V. & Emsley, J. (2015). *J Thromb Haemost* **13**, 580-591.

Renne, T., Schmaier, A. H., Nickel, K. F., Blomback, M. & Maas, C. (2012). *Blood* **120**, 4296-4303.

Reuten, R., Nikodemus, D., Oliveira, M. B., Patel, T. R., Brachvogel, B., Breloy, I., Stetefeld, J. & Koch, M. (2016). *PLoS One* **11**, e0152386.

Towbin, H., Staehelin, T. & Gordon, J. (1979). *Proc Natl Acad Sci U S A* **76**, 4350-4354.

Ullah, H., Scappini, E. L., Moon, A. F., Williams, L. V., Armstrong, D. L. & Pedersen, L. C. (2008). *Protein Sci* **17**, 1771-1780.

Ulmer, J. S., Lindquist, R. N., Dennis, M. S. & Lazarus, R. A. (1995). *FEBS Lett* **365**, 159-163.

Waugh, D. S. (2016). *Postepy Biochem* **62**, 377-382.

Weidmann, H., Heikaus, L., Long, A. T., Naudin, C., Schluter, H. & Renne, T. (2017). *Biochim Biophys Acta* **1864**, 2118-2127.

Winter, G., Waterman, D. G., Parkhurst, J. M., Brewster, A. S., Gildea, R. J., Gerstel, M., Fuentes-Montero, L., Vollmar, M., Michels-Clark, T., Young, I. D., Sauter, N. K. & Evans, G. (2018). *Acta Crystallogr D Struct Biol* **74**, 85-97.

Wong, S. S., Ostergaard, S., Hall, G., Li, C., Williams, P. M., Stennicke, H. & Emsley, J. (2016). *Blood* **127**, 2915-2923.

Worm, M., Kohler, E. C., Panda, R., Long, A., Butler, L. M., Stavrou, E. X., Nickel, K. F., Fuchs, T. A. & Renne, T. (2015). *Ann Transl Med* **3**, 247.

Zhou, D., Lobo, Y. A., Batista, I. F., Marques-Porto, R., Gustchina, A., Oliva, M. L. & Wlodawer, A. (2013). *PLoS One* **8**, e62252.

Supporting information for article:

Crystal structures of the recombinant β-Factor XIIa protease with bound Pro-Arg and Thr-Arg substrate mimetics

Monika Pathak[1*], Rosa Manna[1*], Chan Li[1], Bubacarr G Kaira[1], Badraldin Kareem Hamad[1], Danilo D Belviso[1], Camila R Bonturi[2], Ingrid Dreveny[1], Peter M Fischer[1], Lodewijk V Dekker[1], Maria Luiza V Oliva[2], and Jonas Emsley[1¶].

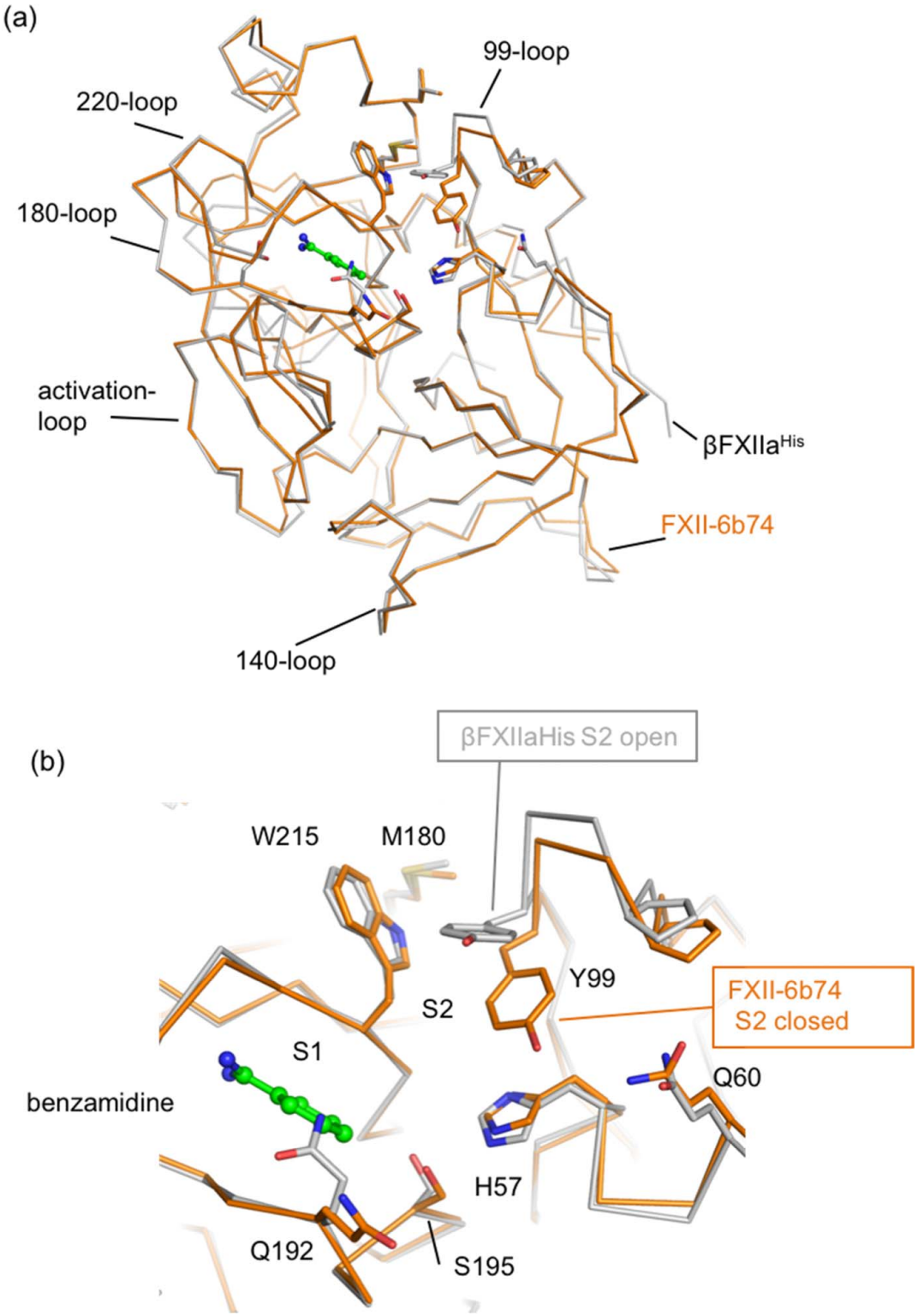

Figure S1. (a) Recombinant and plasma purified βFXIIa crystal structures compared. C-α traces are displayed for superposed recombinant protease domain βFXIIa^{His} (grey) and plasma purified βFXIIa (pdb code:6b74, orange) generated using PyMol and overall r.m.s.d. for the superposition=0.513 Å (b) Zoomed in view illustrating key differences in the region of the 99-loop and the S2 pocket with key residues highlighted as sticks.

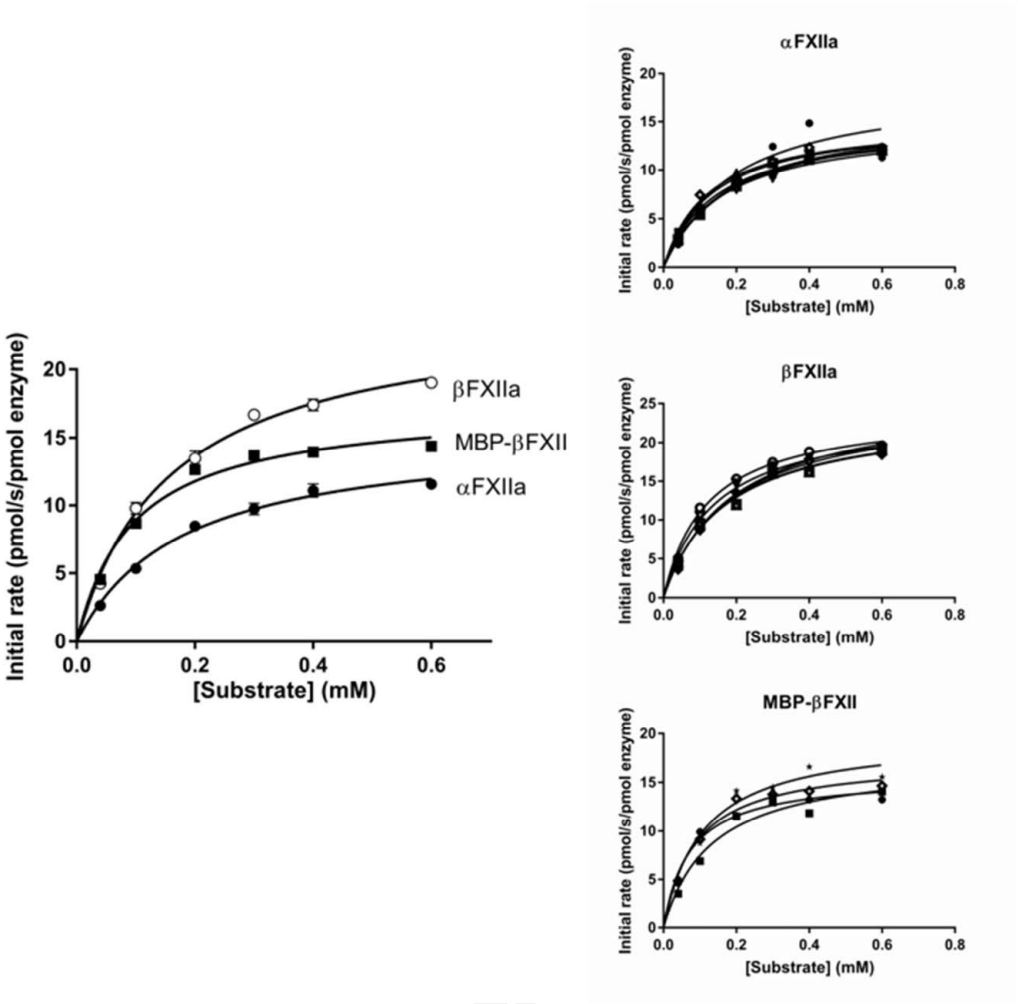

Figure S2. Plot of enzyme substrate S2302 hydrolysis as function of substrate concentration for recombinant MBP-βFXIIa^{His}, commercial βFXIIa and commercial αFXIIa. OD at 405 nm (as a function of time (min)) for different concentrations of S2302 added to βFXIIa was used to calculate initial rate (pmols^{-1}) of S2302 hydrolysis (left panel). An average of four experiments were used to determine the enzyme kinetics with independent plots superposed shown on the right panels.

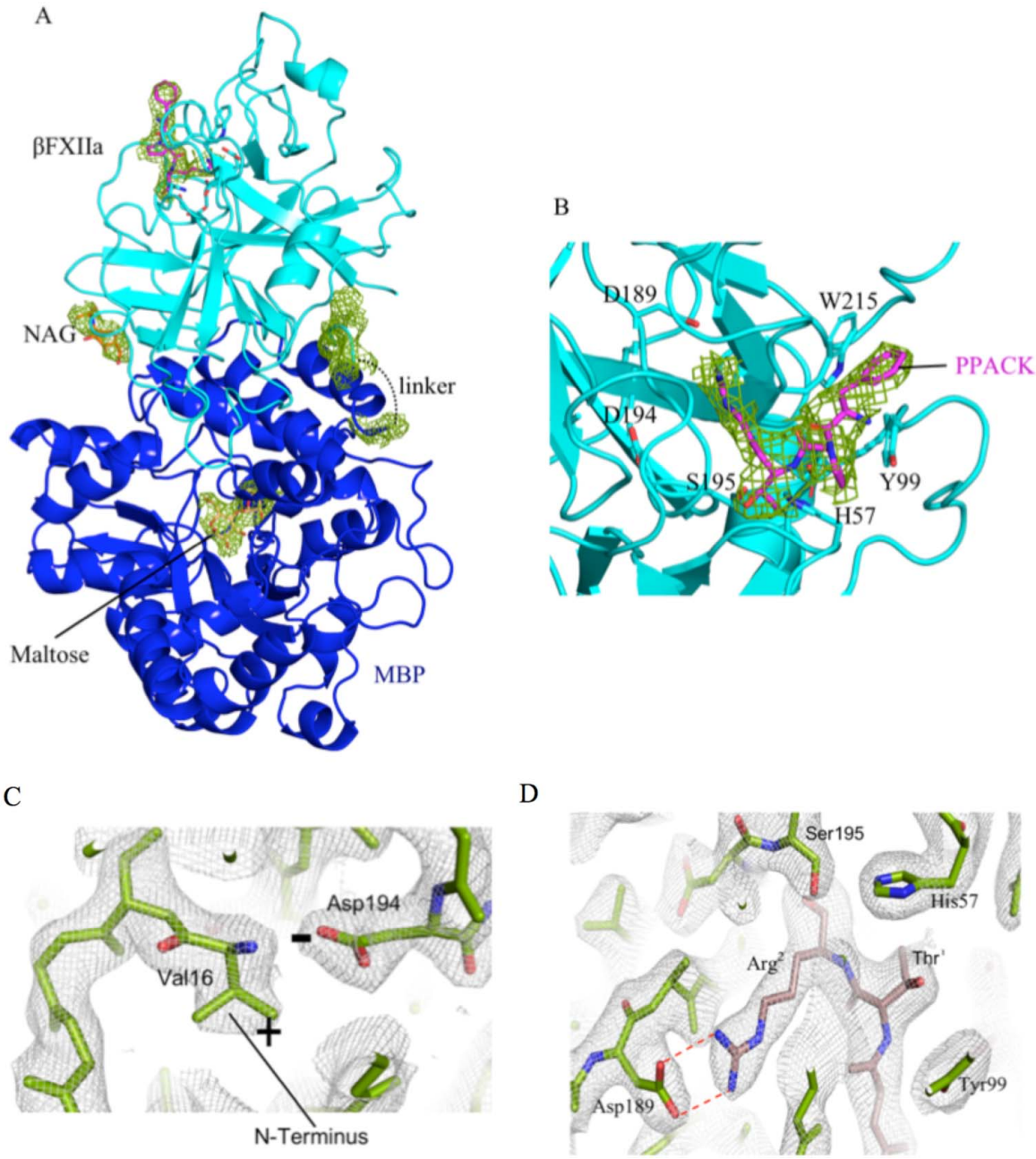

Figure S3. Cartoon diagram of the MBP-βFXIIa^{His} structure. MBP and βFXIIa^{His} are shown in blue and cyan respectively. *A,* Electron density 2Fo-Fc maps at 4 Å resolution are contoured at 1.0 σ are shown (green) for the bound maltose and N-linked glycan. *B*. Electron density Fo–Fc difference map showing the position of PPACK in the βFXIIa^{His} active site cleft. Side chains interacting with PPACK and the catalytic triad are drawn as sticks. *C,* Electron density 2Fo-Fc maps at 2.54 Å resolution are shown contoured at 1.0 σ (grey) for βFXIIa^{His} in the region of the N-terminus and *D* shows the S1 pocket with crystal symmetric residue Arg^2 (brown) forming electrostatic interactions (red dotted lines).

Table S1. Summary of the docked EcTI and βFXIIa interfacial interactions and bond distances.

	βFXIIa [atom]	Dist. [Å]	EcTI [atom]
Hydrogen Bonds			
1	GLN 192[NE2]	2.91	ASN13[O]
2	TYR 99[OH]	2.79	THR 61[OG1]
3	GLY 215[N]	3.18	PRO62[O]
4	GLY 193[N]	2.71	ARG64[O]
7	SER 177[OG]	2.77	HIS114[O]
11	SER 214[O]	3.24	ARG64[N]
17	GLN 60[OE1]	2.83	THR69[OG1]
Salt Bridges			
1	GLU 149[OE2]	3.45	ARG12[NH1]
2	GLU 149[OE1]	2.74	ARG12[NH1]
3	GLU 149[OE2]	2.81	ARG12[NH2]
4	GLU 149[OE1]	2.73	ARG12[NH2]
5	ASP 189[OD1]	2.73	ARG64[NH1]
6	ASP 189[OD2]	3.69	ARG64[NH1]
7	ASP 189[OD1]	3.18	ARG64[NH2]
8	ASP 189[OD2]	2.73	ARG64[NH2]
9	ASP 222[OD2]	3.95	ARG92[NE]
10	ASP 222[OD2]	2.76	ARG92[NH1]
11	GLU 146[OE1]	3.11	ARG92[NH1]
12	GLU 146[OE1]	3.8	ARG92[NH2]
13	GLU 149[OE2]	3.42	ARG174[NH1]

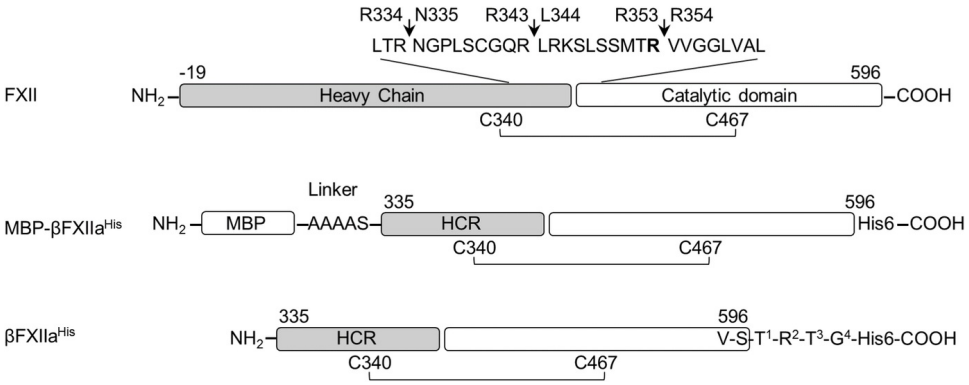
R334 N335 R343 L344 R353 R354
LTR NGPLSCGQR LRKSLSSMTR VVGGLVAL
FXII
-19
596
NH2
Heavy Chain
Catalytic domain
COOH
C340
C467
Linker
335
596
MBP-βFXIIaHis
NH2
MBP
AAAAS
HCR
His6 COOH
C340
C467
335
596
βFXIIaHis
NH2
HCR
V-S-T1-R2-T3-G4-His6-COOH
C340
C467

Figure 1

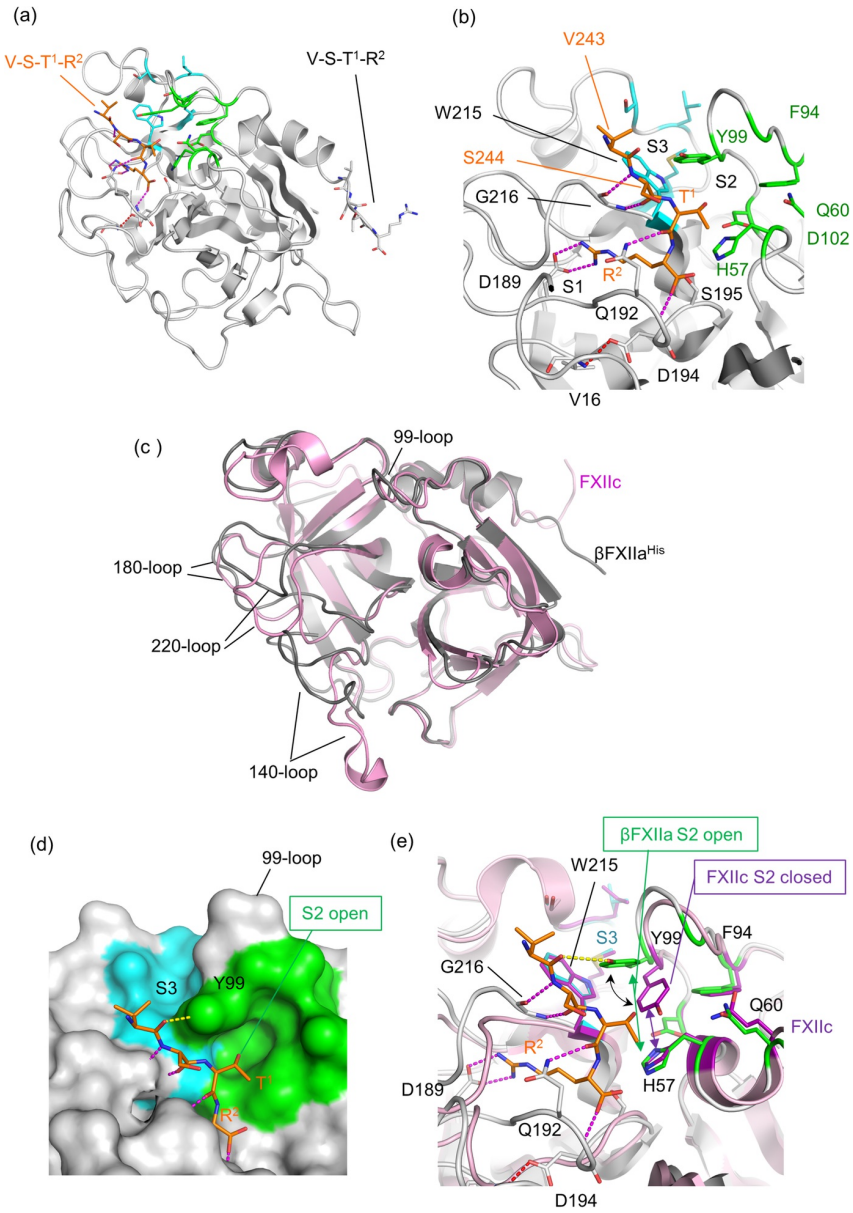
(a)
V-S-T¹-R²
V-S-T¹-R²
(b)
V243
W215
F94
Y99
S3
S244
S2
G216
T¹
Q60
D102
R²
H57
D189
S1
S195
Q192
D194
V16
(c)
99-loop
FXIIc
βFXIIa^His
180-loop
220-loop
140-loop
(d)
99-loop
S2 open
S3
Y99
T¹
R²
(e)
βFXIIa S2 open
W215
FXIIc S2 closed
S3
Y99
F94
G216
Q60
FXIIc
R²
D189
H57
Q192
D194

Figure 2

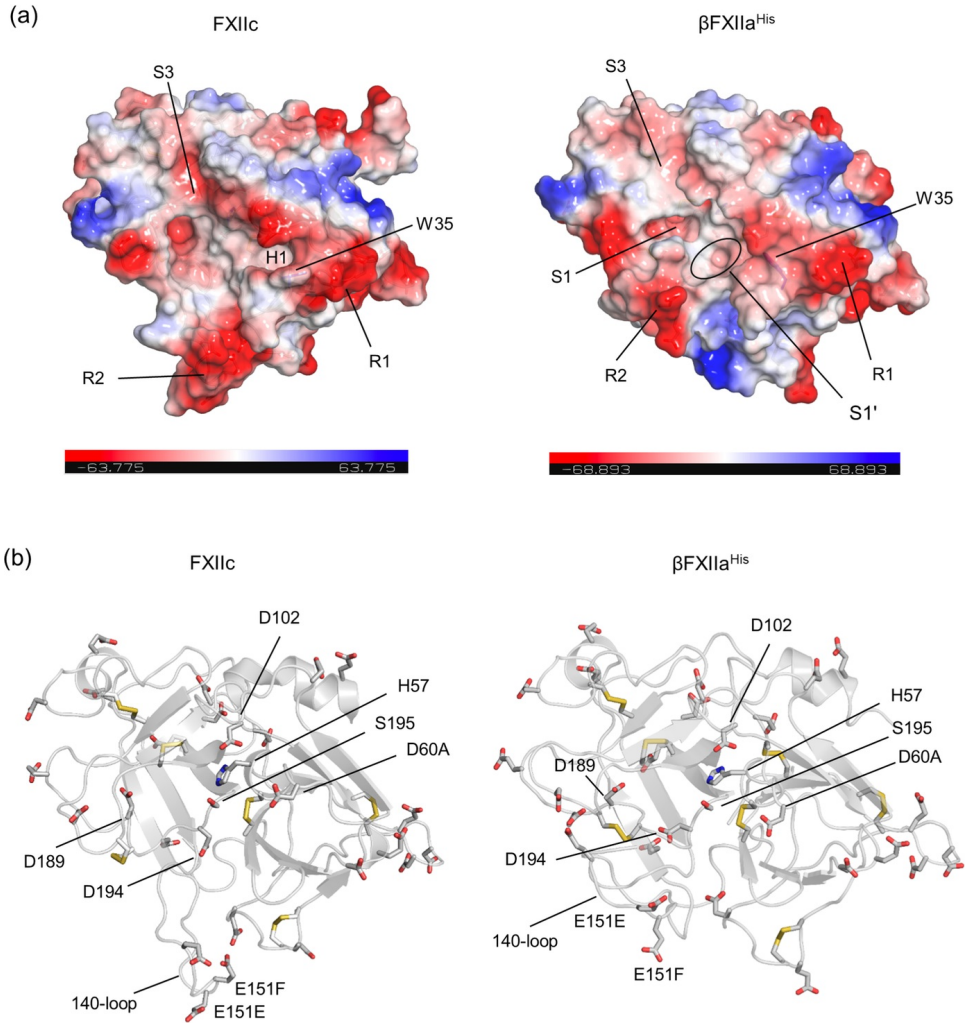
(a)
FXIIc
βFXIIa^His
S3
W35
H1
R1
R2
S1
S1'
-63.775
63.775
-68.893
68.893
(b)
FXIIc
βFXIIa^His
D102
H57
S195
D60A
D189
D194
140-loop
E151F
E151E

Figure 3

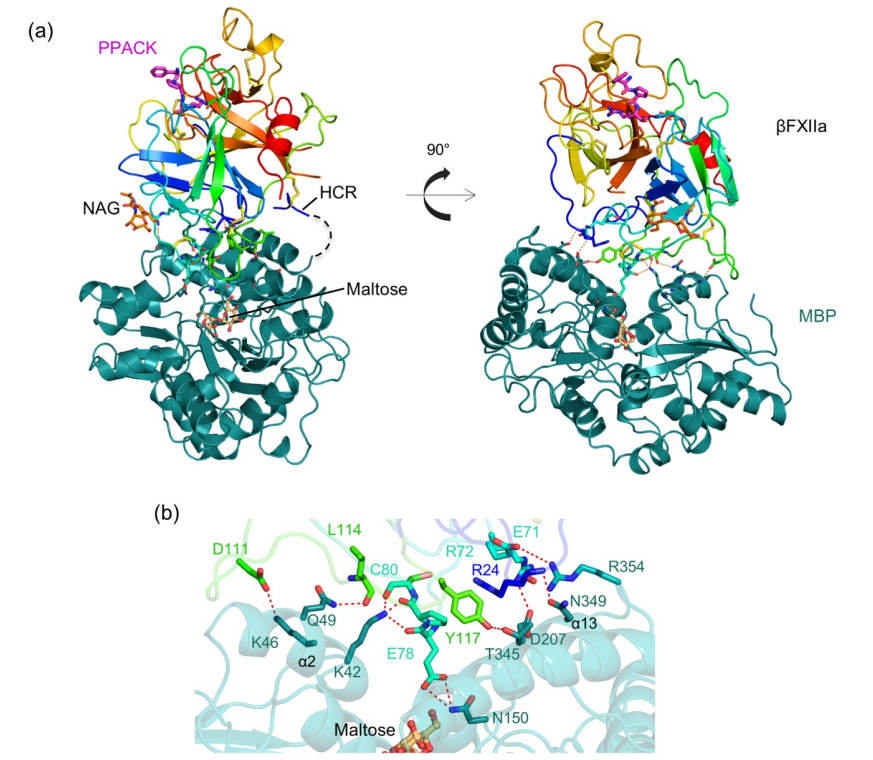
(a)
PPACK
NAG
HCR
Maltose
90°
βFXIIa
MBP
(b)
D111
L114
C80
Q49
K46
α2
K42
E78
Y117
R72
E71
R24
R354
N349
α13
D207
T345
N150
Maltose
(c)
S175
L176
F94
W215
S3
Y99
PPACK
S2
Q60
H57
D189
S195
D194
V16
(d)
S3
S2
D60A
S1
H1
closed
oxyanion
hole
E151E
S1'

Figure 4

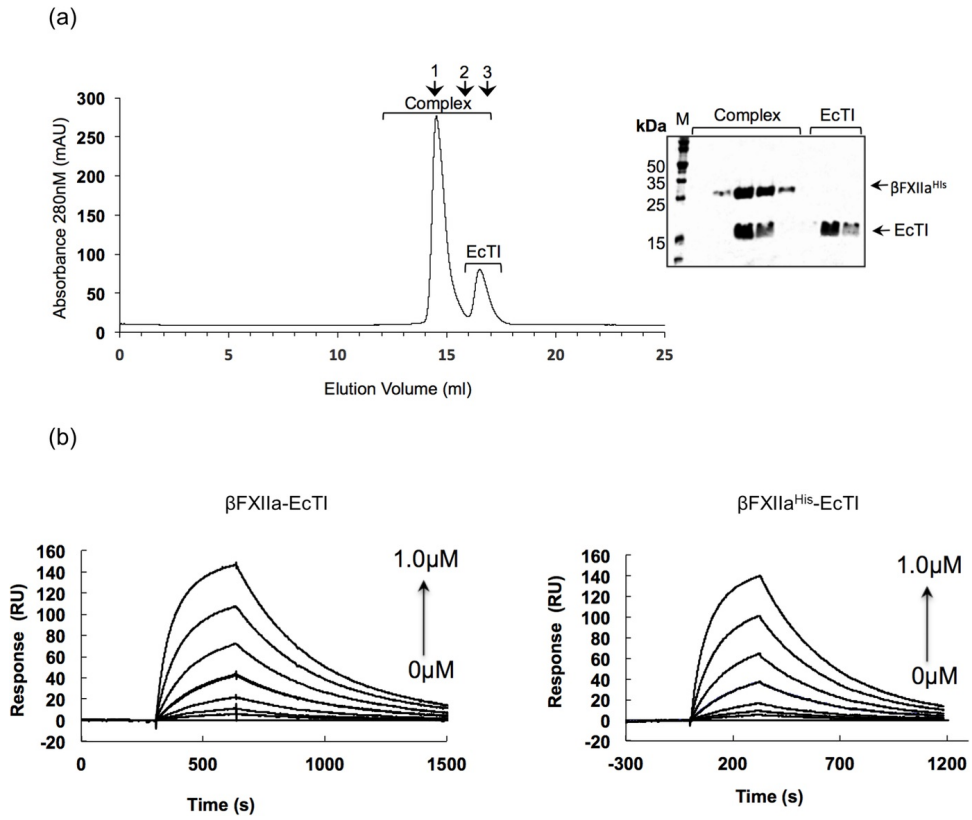
(a)
1 2 3
Complex
EcTI
Absorbance 280nM (mAU)
Elution Volume (ml)
kDa
M
Complex
EcTI
βFXIIa^His
EcTI
(b)
βFXIIa-EcTI
βFXIIa^His-EcTI
Response (RU)
Time (s)
1.0μM
0μM

Figure 5

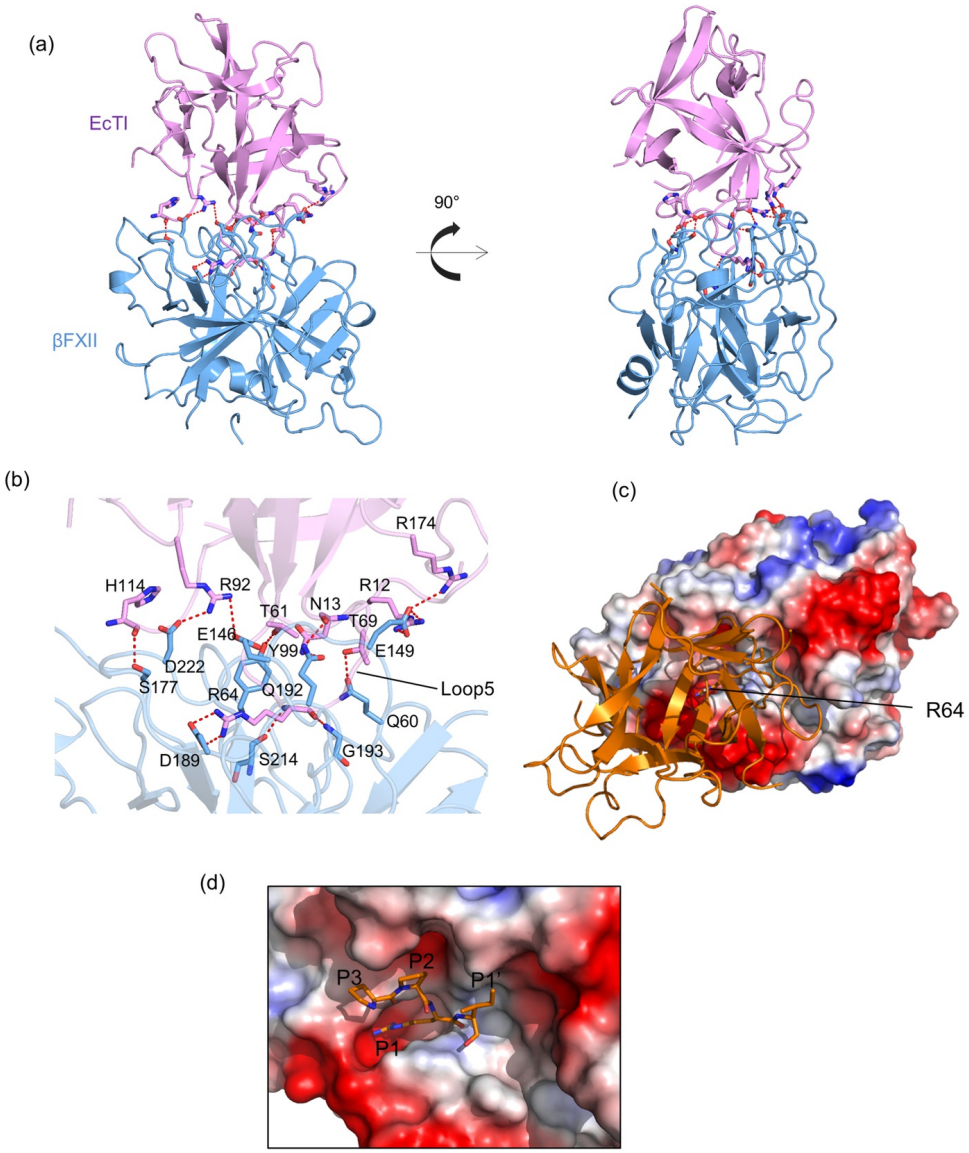
(a)
EcTI
βFXII
90°
(b)
R174
H114
R92
T61
N13
R12
T69
E146
Y99
E149
D222
S177
R64
Q192
Loop5
Q60
D189
S214
G193
(c)
R64
(d)
P3
P2
P1'
P1

Figure 6

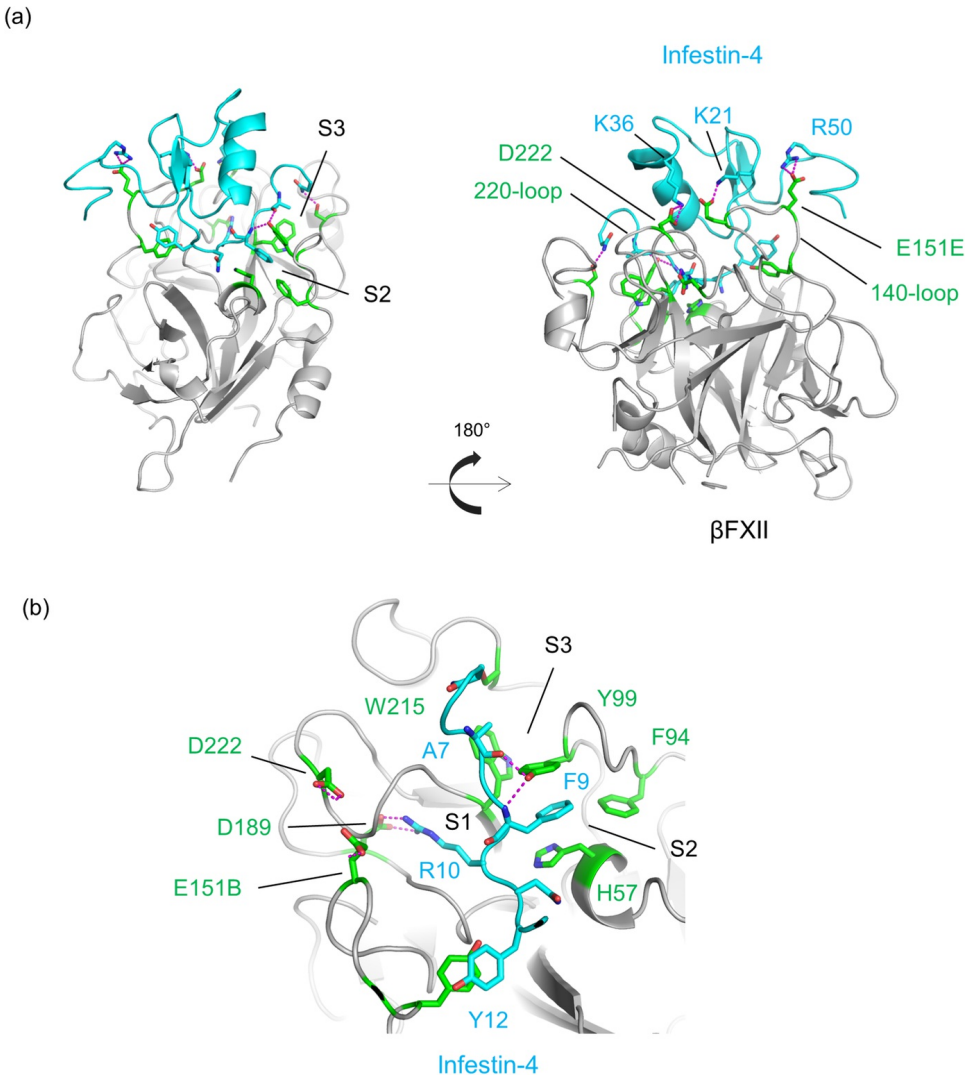
(a)
Infestin-4
S3
S2
K36
K21
R50
D222
220-loop
E151E
140-loop
180°
βFXII
(b)
S3
W215
Y99
D222
A7
F94
F9
D189
S1
S2
R10
E151B
H57
Y12
Infestin-4

Figure 7