doi:10.1016/S0022-2836(02)00559-4 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (2002) 321, 57–68

JMB

Crystal Structure of Earthworm Fibrinolytic Enzyme Component A: Revealing the Structural Determinants of its Dual Fibrinolytic Activity

Yong Tang, Dongcai Liang, Tao Jiang, Jiping Zhang, Lulu Gui and Wenrui Chang*

National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101 People's Republic of China

**Corresponding author*

Earthworm fibrinolytic enzyme component A (EFEa) from *Eisenia fetida* is a strong fibrinolytic enzyme that not only directly degrades fibrin, but also activates plasminogen. Proteolytic assays further revealed that it cleaved behind various P1 residue types. The crystal structure of EFEa was determined using the MIR method and refined to 2.3 Å resolution. The enzyme, showing the overall polypeptide fold of chymotrypsin-like serine proteases, possesses essential S1 specificity determinants characteristic of elastase. However, the β strand at the west rim of the S1 specificity pocket is significantly elongated by a unique four-residue insertion (Ser-Ser-Gly-Leu) after Val217, which not only provides additional substrate hydrogen binding sites for distal P residues, but also causes extension of the S1 pocket at the south rim. The S2 subsite of the enzyme was partially occluded by the bulky side-chain of residue Tyr99. Structure-based inhibitor modeling demonstrated that EFEa's S1 specificity pocket was preferable for elastase-specific small hydrophobic P1 residues, while its accommodation of long and/or bulky P1 residues was also feasible if enhanced binding of the substrate and induced fit of the S1 pocket were achieved. EFEa is thereby endowed with relatively broad substrate specificity, including the dual fibrinolysis. The presence of Tyr99 at the S2 subsite indicates a preference for P2-Gly, while an induced fit of Tyr99 was also suggested for accommodation of bigger P2 residues. This structure is the first reported for an earthworm fibrinolytic enzyme component and serine protease originating from annelid worms.

Keywords: crystal structure; earthworm fibrinolytic enzyme; *Eisenia fetida*; serine protease; substrate specificity

Introduction

Thrombosis is one of the most widely occurring diseases in modern life, which often causes disability and death. Fibrinolytic enzymes degrade fibrin, the major protein component of blood clots. Medications using fibrinolytic enzymes are the most effective methods used in the treatment of thrombosis. A variety of fibrinolytic enzymes, such as tPA, uPA, and bacterial plasminogen activator streptokinase, have been extensively studied and used as thrombolytic agents. However, these agents are expensive and suffer from a number of significant limitations, such as fast clearance, reocclusion and bleeding complications.[1] Therefore, novel thrombolytic agents with better fibrinolytic efficiency and less adverse side effects are eagerly sought.

Present address: Y. Tang, Department of Cancer Biology, Dana-Farber Cancer Institute, 44 Binney Street, Boston, MA 02115, USA.

Abbreviations used: DFP, diisopropyl fluorophosphate; EFE, earthworm fibrinolytic enzyme; EFEa, EFE component A; HLE, human leukocyte elastase; MIR, multiple isomorphous replacement; PPE, porcine pancreatic elastase; SBTI, soybean trypsin inhibitor; tPA, tissue-type plasminogen activator; TLCK, tosyl-L-lysine chloromethyl ketone; uPA, urokinase-type plasminogen activator; VR, variable region; ε-ACA, ε-aminocaproic acid.

E-mail address of the corresponding author: wrchang@sun5.ibp.ac.cn

0022-2836/02/$ - see front matter

Earthworms have been used as thrombolytic agents in Chinese medicine in East Asia for several thousand years. Only recently, earthworm fibrinolytic enzymes (EFEs) from *Lumbricus rubellus* and *Eisenia fetida* were characterized[2,3] and became commercially available in Korea and China as novel oral-administered fibrinolytic agents for prevention and treatment of cardiac and cerebrovascular diseases.[4,5] EFE is stable for long-term storage at room temperature and its oral administration is very convenient. Earthworms, the raw material of EFE, can be easily raised, making EFE a relatively inexpensive thrombolytic agent suitable for large-scale production.

The therapeutic value of EFE has made more detailed study on this enzyme highly desirable. Further purification and characterization revealed that EFE from *L. rubellus* was composed of six serine protease components all having considerable albeit different *in vitro* fibrinolytic activities.[6] The substrate/inhibitor specificity of these components also differed, with two components (F-III-1 and F-III-2) characterized as trypsin-like and three (F-I-0, F-I-1, and F-I-2) as chymotrypsin-like, with the other component (F-II) still unclear.[7] Among these components, F-II is very interesting for its extraordinary substrate specificity and inhibition characteristics. F-II showed relatively higher fibrinolytic and caseinolytic activities than other fibrinolytic components.[6] Moreover, F-II cleaved several peptide bonds of oxidized insulin B chains and β-amyloid 1–40 behind basic, aromatic or hydrophobic residues.[8] However, in striking contrast, F-II was almost inactive with synthetic chromogenic substrates specific for trypsin, chymotrypsin or elastase.[6] F-II was completely inhibited by DFP and significantly inhibited by soybean trypsin inhibitor (SBTI) and aprotinin, while its inhibition by TLCK, ε-ACA, or elastatinal was very limited.[7] The reasons for all these observations have not yet been identified.

Recently, we purified and crystallized a novel earthworm fibrinolytic enzyme from *E. fetida* which we named earthworm fibrinolytic enzyme component A (EFEa).[9] This enzyme exhibits higher fibrinolytic effects in plasminogen-rich fibrin plates than in plasminogen-free plates, indicating that it acts not only as a direct fibrinolytic enzyme, but also as a plasminogen activator. The dual fibrinolytic activity of EFEa has been further confirmed by its cleavage of bovine plasminogen (Li & He, unpublished results). In assay, EFEa cleaved plasminogen at five peptide bonds, three of the P1 residues being hydrophobic and the other two being basic (Pn and Pn' denote the nth amino acid at the N and C-terminals of the cleaved peptide bond; Sn and Sn' denote the corresponding binding sites of the protease[10]). The cleavage at the peptide bond R557-I558 (plasminogen numbering) led to the generation of active plasmin.

The EFEa enzyme, purified by a method similar to that for F-II,[6] has the same first 24 N-terminal amino acid residues and essentially the same relative molecular mass (24,667 *versus* 24,664, mass spectroscopy results) as F-II. The unique incomplete fibrinolysis of EFEa as observed in the fibrin plate assay also resembles that of F-II.[7] Further cDNA sequencing of EFEa (GenBank, access number AF393512) revealed that the cDNA-deduced sequence of EFEa with 242 residues was almost identical to that of the earthworm fibrinolytic enzyme from *Lumbricus bimastus* (GenBank, AF109648), the only difference being residue 219, which is Ala in EFEa and Thr in the latter (chymotrypsinogen numbering[11] is used throughout this paper unless otherwise indicated). The similarity of these sequences and the close evolutionary relationship of the three earthworm species† enable us to infer that EFEa from *E. fetida* is extremely homologous, if not identical, to F-II from *L. rubellus*.

Sequence alignment (Figure 1) showed that EFEa shares moderate sequence identity with its closely related serine proteases with known structures: 36% to porcine pancreatic elastase[12] (PPE; PDB code: 4EST), 28% to human leukocyte elastase[13] (HLE; 1PPG), 35% to trypsin[14] (3PTB), 34% to chymotrypsin[15] (1DLK), and 32% to the catalytic domain of uPA[16] (1LMW), tPA[17] (1RTF), or plasmin[18] (1BML). The sequence identities of EFEa to other earthworm serine proteases are also not significant: about 35% to F-III-1 and F-III-2[19] and 31% to earthworm chymotrypsinogen (GenBank, CAA11132). The highest sequence homology of EFEa is observed with lugworm (*Arenicola marina*) chymotrypsinogen (GenBank, CAA64472), with an identity of 44%.

The crystal structure of the EFEa from *E. fetida* was solved using the multiple isomorphous replacement (MIR) method to study the mechanism underlying EFEa's dual fibrinolytic activities and to locate the structural determinants of its extraordinary substrate specificity. A detailed description of the EFEa structure is given here with emphasis on the novel structural basis of its substrate specificity. Structure comparisons were given with related proteins, especially with PPE. The three-dimensional structure presented here is the first reported for an earthworm fibrinolytic enzyme component and serine protease originating from annelid worms. It provides a template for accurately modeling other homologous family members.

Results

Overall structure

The EFEa structure has been refined to an R value of 0.191 (R free = 0.236) at 2.3 Å resolution. EFEa is roughly spherical with a radius of about 25 Å. The overall structure is folded into two six-stranded barrel-like sub-domains connected by

† See http://www.ncbi.nlm.nih.gov for the taxonomy.

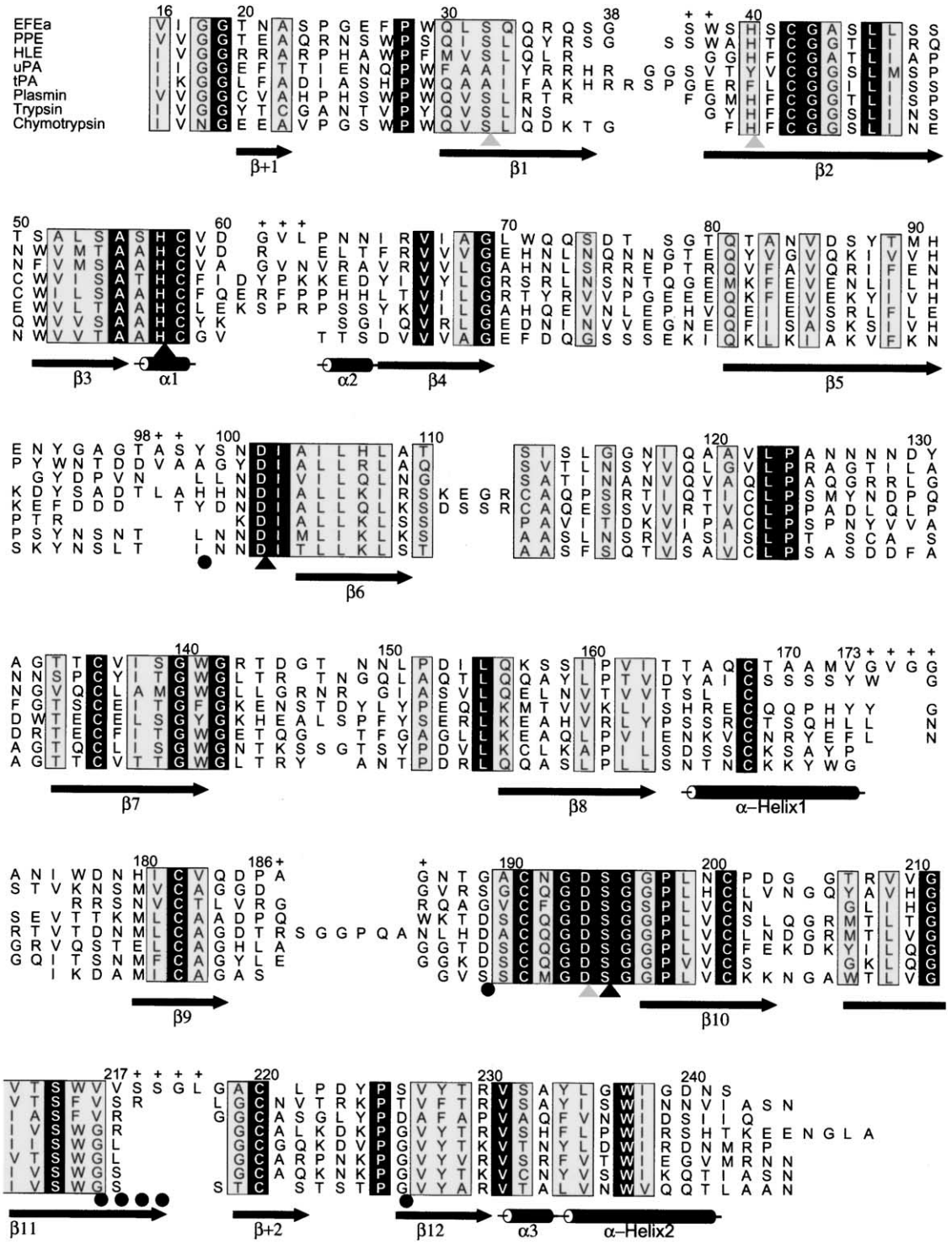

Figure 1. Sequence alignment of EFEa with closely related serine proteases. Chymotrypsinogen numbering is given on the top and the inserted residues are marked with +. The secondary structures are shown as barrels for helices (α for 3_{10} helix and α-helix for standard 3.6_{13} helix) and arrows for β strands. Catalytic and zymogen triads are marked with black and gray triangles, respectively, while other important residues are marked with black circles.

three trans-domain straps, showing the typical fold of a chymotrypsin-like serine protease (Figure 2). The active-site cleft and the catalytic residues are located at the junction of both barrels, with the active site cleft perpendicular to the junction. Inter-segment hydrogen bonds result in the formation of two additional β strands, i.e. the N terminal β strand Thr20-Ala22 (β + 1) and the β strand Ala219-Leu221 (β + 2) located at the south boundary of the S1 specificity pocket (Figure 2). The surface is made up of various loops, turns, and two α-helices spanning residues

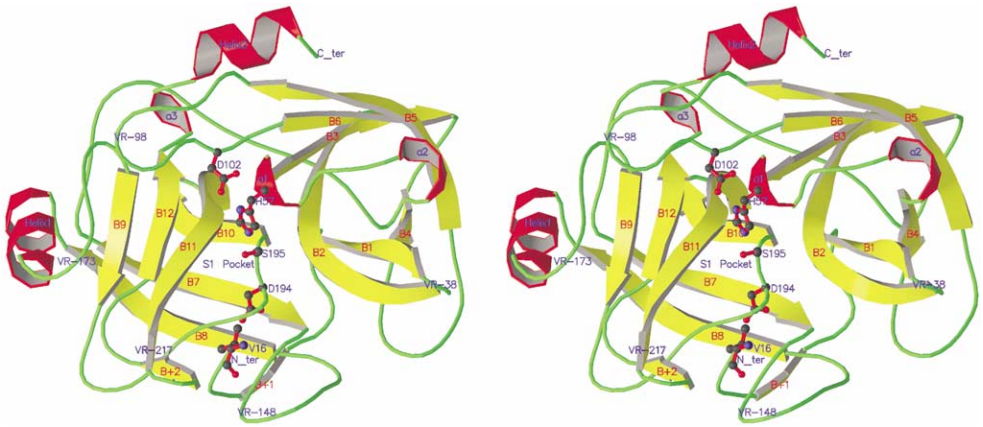

Figure 2. Ribbon representation of EFEa structure in standard orientation. β Strands are shown in yellow, helices in red, and loop regions in green. The secondary elements are labeled as in Figure 1 and the variable regions that would be involved in substrate binding are labeled with VR prefixes and the positions where the insertion or deletion occurs. Catalytic triad residues His57, Asp102 and Ser195, and Val16 and Asp194 that form the very important internal salt bridge, are also shown (in ball-and-stick). The S1 specificity pocket site is also noted.

Thr165-Val173 and residues Tyr234-Asn241 (Figures 1 and 2).

As in almost all other active chymotrypsin-like serine proteases, the α-amino group of the EFEa N-terminal residue (Val16) forms an internal salt bridge with the side-chain carboxyl group of Asp194 (Figure 2), which generates an oxyanion hole,[20] suggesting that the structural changes leading to chymotrypsinogen and trypsinogen activations[21,22] also occur in the proteolytic generation of active EFEa. Furthermore, the presence of the "zymogen triad" residues,[23] namely Ser32, His40, and Asp194, in EFEa also suggests that the zymogen form of EFEa will exhibit a high level of zymogenicity. The C-terminal residue of EFEa, which is well defined in the electron density map, is Ser242. The crystal packing does not appear to allow space for an additional residue, Arg243 in the cDNA-deduced sequence. Subsequent mass spectroscopy results further confirmed the absence of Arg243 in the EFEa enzyme.[9] The removal of Arg243 may result from post-translational trimming, as was also proposed for the deleted C-terminal residues of HLE.[24] The electron density of Ala219, the only residue different from that in the earthworm fibrinolytic enzyme from *L. bimastus*, was unambiguously defined in our final $2F_o - F_c$ and SA omit maps. Unlike other earthworm serine proteases with known sequences, EFEa has merely eight cysteine residues that form the four disulfide bridges conserved among most serine proteases.[25] Having 13 aspartate and two glutamate residues but only one lysine and five arginine residues, EFEa is a fairly acidic molecule, with a pI value of 4.2. Apart from Asp102 and Asp194, which are involved in the catalytic triad and in the internal salt bridge formation, respectively, all charged side-chains are exposed to the bulk solvent. Only three charged residues are located in the vicinity of the active site, with Arg143 to the south, Asp60 to the north, and Arg35 to the east of the S1 specificity pocket (Figure 3(a)). Most other charged residues are located on the rear side of the enzyme.

The three independent molecules in the asymmetric unit of the crystal structure (named A, B and C) exhibit high levels of structural conservation. Optimal superposition of the 241 C^{α} carbon atom pairs of molecules A and B, A and C, and B and C in the asymmetric unit using a 0.5 Å distance cutoff results in 230, 215, and 212 topologically equivalent atoms with root mean square (RMS) deviations of 0.21, 0.22, and 0.22 Å, respectively.

Optimal superposition of EFEa with PPE (PDB ID: 4EST), HLE (1PPG), uPA (1LMW), tPA (1RTF), plasmin (1BML), trypsin (3PTB) and chymotrypsin (1DLK) resulted in 199, 186, 200, 202, 196, 201 and 199 topologically equivalent C^{α} atoms with RMS deviations of 0.81, 0.98, 1.12, 1.10, 0.99, 0.93 and 1.01 Å, respectively, when only the C^{α} carbon atom pairs within 2.5 Å were included. The data indicate the topological homology of these enzymes with EFEa: porcine pancreatic elastase is the most similar, followed by trypsin and others. This order is in agreement with that of the sequence identity as described above. These topological equivalences form the basis for the alignment of the eight serine proteases as shown in Figure 1, with the EFEa numbering based on its equivalence with chymotrypsin. A proper alignment of EFEa with chymotrypsin requires single-residue deletions at chymotrypsin positions 77, 148, and 205, and residue insertions after chymotrypsin positions 38, 60, 98, 173, 186, and 217 (see Figure 1). These deletions and insertions, all occurring in the surface regions, constitute the variable regions[25] (VRs) characteristic of EFEa. The variable regions that were likely to contribute to substrate interaction are indicated in Figure 2 and are described below.

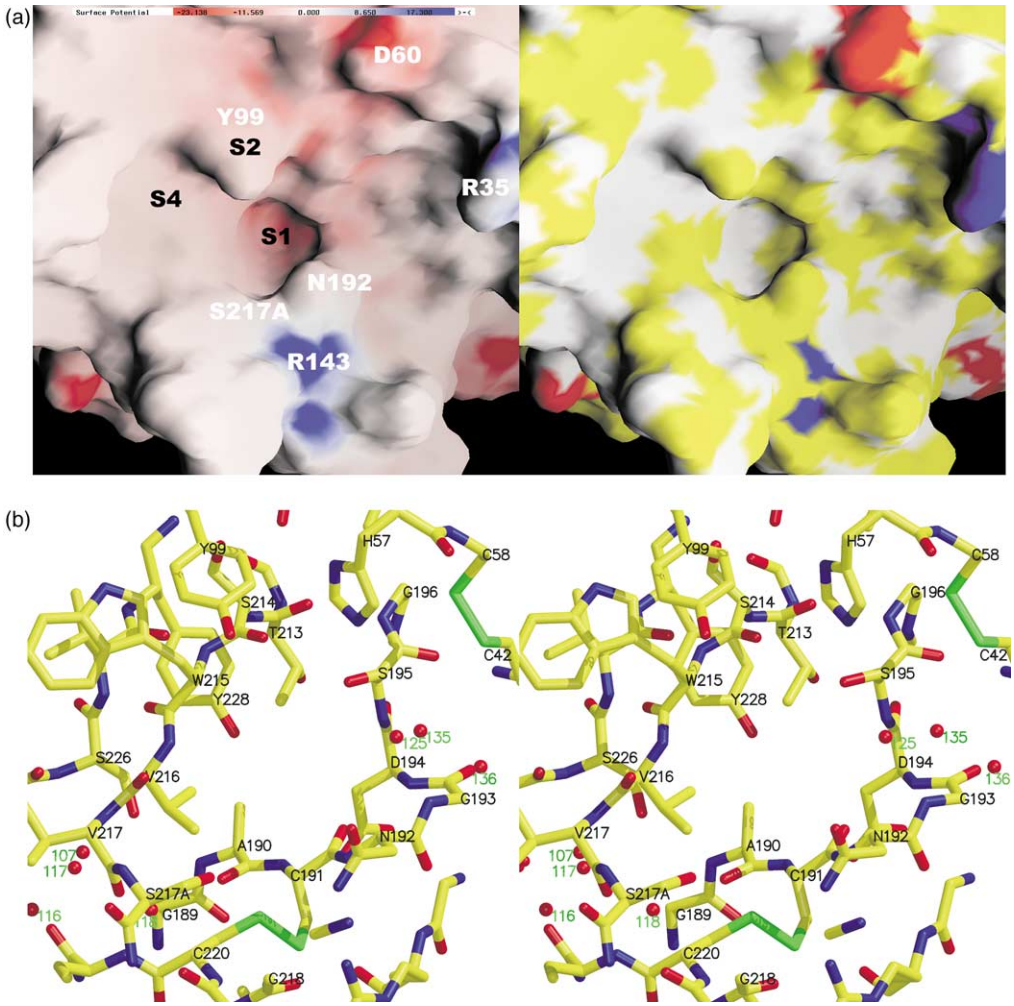

Figure 3. The molecular surface of EFEa with (a) electrostatic potential distribution and (b) hydrophobicity property. In (a), the regions with negative and positive charges are shown in red and blue, respectively. The S1, S2 (Tyr99) and S4 subsites are marked, as are the charged residues Asp60, Arg35 and Arg143 near the S1 pocket and residues Ser217A and Asn192 at the south border of the S1 pocket. In (b), the hydrophobic surface is shown in white, the polar surface in yellow, and the charged surface in blue (positive) and red (negative). (c) The EFEa active-site residue distribution. A sphere with 11 Å radius was drawn centered at the S1 specificity pocket. One-letter codes were used for the residues with the solvent molecules drawn as red balls and labeled with green numerals.

Substrate binding subsites

The specificity of the three subclasses of the chymotrypsin family, namely chymotrypsin, trypsin, and elastase, is usually most pronounced at the S1 substrate specificity pocket, where they can be readily defined by inspecting a small number of crucial amino acid residues.[26] The solution of the EFEa structure makes a definite classification of EFEa possible. In EFEa, the S1 pocket was walled by residue His57 (in the northeast), segments Pro225-Val227 (in the middle west), Thr213-Cys220 (from the lower north to the upper south), and Gly189-Ser195 (from the lower south to the upper east), and closed on the bottom by Tyr228 (Figure 3(c)). The backbones of these residues, except those in segment Val217-Cys220, are readily superimposable with their topological equivalences in other closely related proteases, especially PPE (Figure 4). The topological conservation of these residues suggests that the specificity of EFEa will be determined by the residue type of the primary specificity determinants in the S1 pocket, namely residues 189, 216, and 226, which are Gly189, Val216, and Ser226 in EFEa. These residues, being very characteristic of elastases, substantially define the elastase nature of EFEa. Indeed, the side-chains of Val216 and Ser226 completely block access of a P1 residue with long and/or bulky side-chain towards the direction of Gly189, as demonstrated in Figure 4(a) and (b). At the bottom, the cavity is further restricted by the side-chains of Ala190 and Thr213 before ending at the aromatic ring of Tyr228 (Figure 3(c)). The resultant bowl-shaped geometry, together with the hydrophobic nature of the side-chains of Val216 and Ala190, the disulfide bridge Cys191-Cys220, and the hydrophobic part of Thr213 (Figure 3(b)),

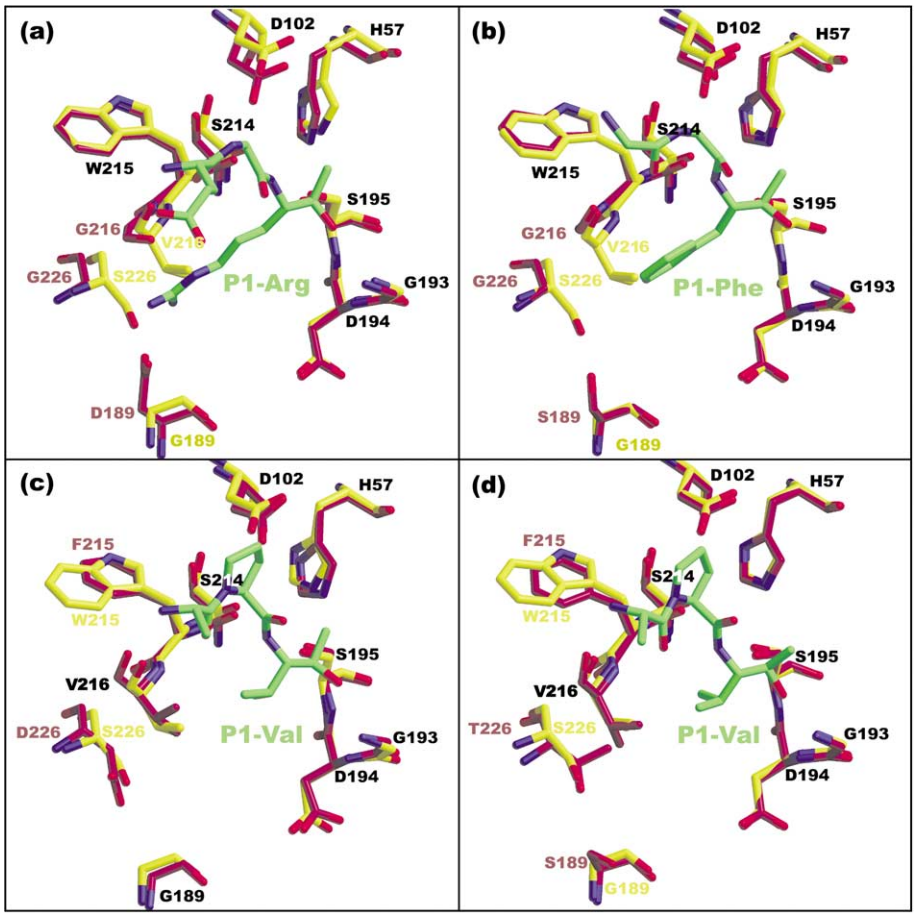

Figure 4. Structural comparison of the active sites of EFEa and its closely related serine proteases based on their global C^{α} superposition. The superposition of some important active site residues is shown, including the catalytic triad and the substrate-discriminating residues (189, 226 and 216) in the S1 specificity pockets of EFEa (in yellow) and (a) tPA[50] (PDB code: 1BDA); (b) chymotrypsin (1DLK); (c) HLE (1PPG) and (d) PPE (4EST) (all in maroon). The specific inhibitors (in green) bound to the various enzymes were introduced into the active site of EFEa to demonstrate the fitting of the P1 residues. Identical residues are labeled in black, while different ones are labeled with their corresponding colors. For clarity, only the P1–P3 residues are drawn for the inhibitors.

make the S1 pocket well adapted to accommodate small hydrophobic residues (Figure 4(c) and (d)).

However, the EFEa S1 pocket was significantly different from other proteases in segment Val216-Cys220 (named VR-217) at the southwest boundary (Figure 5). Chymotrypsin-like and trypsin-like proteases have no insertion in this region, with residue 218 deleted in the trypsin-like proteases. With this residue layout, Gly216 terminates the β strand (β11), with the following residues "going down" towards the conserved Cys220. In PPE, this topology is essentially unaltered by the single-residue insertion of Arg217A, where the protruding side-chain of Arg217A would apparently interfere with substrate binding.[12] In contrast, the VR-217 of EFEa has four inserted residues (Ser-Ser-Gly-Leu) after Val217, which are unique among all the elastases with known sequences. These inserted residues have adapted such an orientation that they effectively elongate the β11 strand to Ser217B and thereby create several additional hydrogen bonding sites for enhanced substrate binding. Gly217C helps the peptide to make a turn around Leu217D, with segment Ala219-Leu221 forming a short additional β strand (β + 2) (Figure 2). The unique conformation of VR-217 allows the S1 pocket to be extended at the south rim so that it can accommodate a large and/or bulky S1 residue.

To the north of the S1 pocket, loop Glu92-Ile103 (named VR-98), though having an unusual reclining conformation, presents a Tyr99 to the S2 subsite (Figure 3). The bulky aromatic side-chain of Tyr99, having a conformation similar to that in tPA,[17] partially blocks the S2 subsite, making it preferable only for a P2-Gly. In contrast, in most other related serine proteases, especially elastases, residue 99 is hydrophobic and of medium size (Figure 1), leaving space for more bulky P2 residues.

To the west of the side-chain of Tyr99, two inserted residues, Gly173C and Gly173D, are presented to the active site surface by segment Thr164-Asn179 (named VR-173), which, together with Ser98B, Tyr99 and Trp215, define a relatively spacious S4 subsite (Figure 3).

As in PPE, the S1′ is unrestrictive and can accommodate almost all residue types. The S2′ subsite of

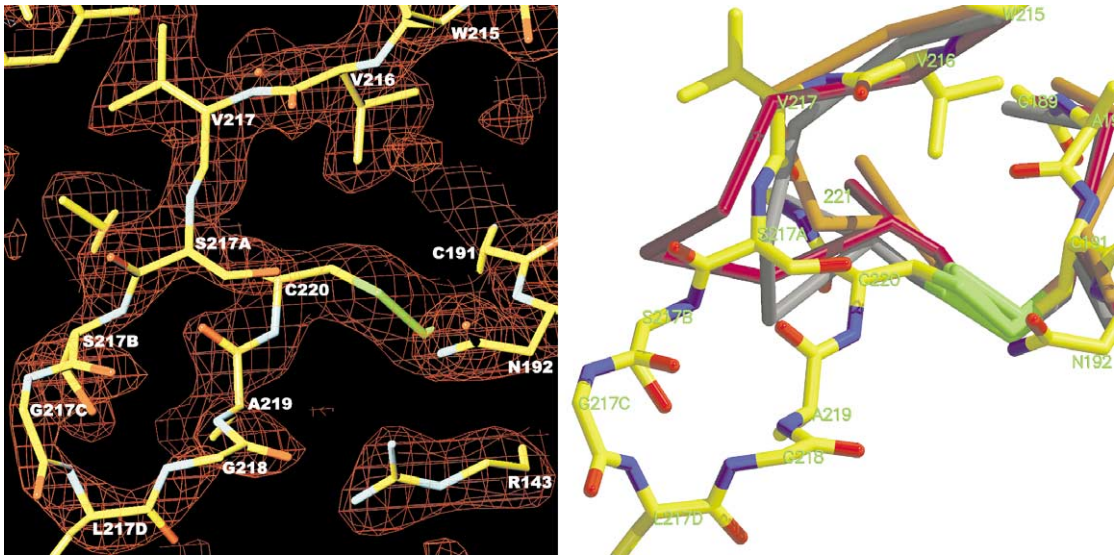

Figure 5. (a) The $2F_o - F_c$ electron density map (in red) of segment Trp215-Cys220 superimposed with the final model. Most of the residues were well defined and the mobility around Leu217D can be observed. (b) Structural comparison of segment 215–220 of EFEa and its closely related serine proteases based on their global C^{α} superposition. The full peptide chain of EFEa is shown, while chymotrypsin (chocolate), trypsin (gray), and PPE (maroon) are shown only in the C^{α} traces. The C220–C191 disulfide bridges in all the enzymes are drawn to demonstrate the spatial conservation.

EFEa, which is defined by the side-chain of Leu151 from loop Gly142-Leu155 (VR-148) and therefore hydrophobic, also resembles that of PPE. The S3′ subsite of EFEa, as defined by the protruding side-chain of Arg35, is also unrestrictive though it differs from Tyr35 in PPE. Therefore, the S′ subsites seem to be of little significance in substrate binding. A loop Gly38-Ser38A (named VR-38) is located outside the east surface of the S1 pocket. As in PPE, VR-38 contains a moderate two-residue insertion integrated in the β2 strand, with the side-chain of Trp38B hidden behind the loop. Therefore, this loop seems less likely to play an important role in the interaction with a substrate, unlike that in tPA, where the positively-charged inserted residues are responsible for PAI-1 binding and fibrin recognition.[27–29]

Discussion

Proposed enhanced substrate binding and induced-fit mechanism

Natural serine protease substrates bind to the active-site cleft of their cognate proteases by building an antiparallel β-strand with residues 214–216.[30] Peptide-derived inhibitors bound in such a canonical mode to serine proteases that are closely related with EFEa have also been reported (see Figure 4). An eight-residue inhibitor based on the Ala-Pro-Val (P1-P3) moiety of the PPE-bound inhibitor[12] (Figure 4(d)) was constructed and modeled into the EFEa active site (Figure 6). This model, having a P1-Arg, a P2-Pro and six alanine residues from P3 to P8, forms a rational antiparallel β-ladder with the extended β11 strand of EFEa. The main chain hydrogen bonds between the inhibitor and EFEa are formed between atoms P1-Arg O and both G193 N and Ser195 N, P2-Pro N and Trp215 O, P3-Ala O and Val216 N, P5-Ala O and Ser217A N, P5-Ala N and Ser217A O, and P7-Ala O and Gly217C N (Figure 6). Therefore, compared with a typical serine protease, EFEa can provide three additional hydrogen bonding sites along the extended β11 strand for binding with the distal P5–P7 residues of a substrate, with minimum restriction to their side-chains. The additional binding energy from the enhanced substrate binding will be essential for the induced fit of the S1 pocket and/or the conformational change of Tyr99 when necessary, as discussed below.

Since the long P1-Arg side-chain could not fit into the S1 pocket in a straightforward manner, another orientation was adapted. Testing of various rotamers revealed that only a "bent" conformation would allow the side-chain to fit into the S1 pocket. In the model, the side-chain of P1-Arg is "sitting" on the upper part of the S1 pocket and is stabilized by hydrogen bonds between its $N^{\eta 1}$ atom and the N192 $O^{\delta 2}$ atom, and between its $N^{\eta 2}$ atom and the Ser217A O^{γ} atom. With such a conformation, aromatic residues such as Tyr and Phe may also be accommodated into the S1 pocket. However, such accommodations would be guaranteed only with a moderate but effective induced fit of the enzyme, specifically a shift of the side-chains of residues Ser217A and/or Asn192 of about 1 Å. An induced fit of Ser217A seems possible since the fragment VR-217 is exposed to the bulk solvent and is inherently mobile around Leu217D as shown by the relatively poor electron density

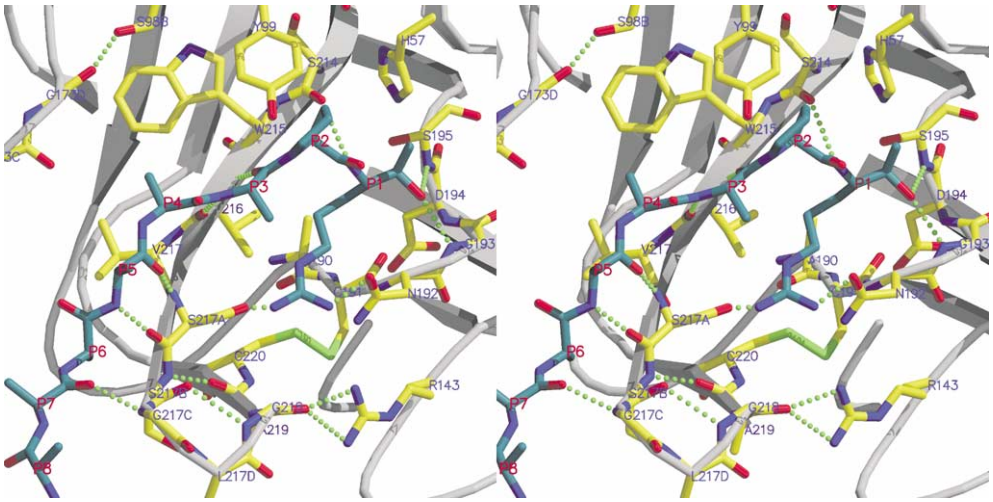

Figure 6. Modeling of a hypothetical eight-residue inhibitor onto the EFEa active site. The enzyme is shown as ribbons (in light gray), with important residues involved in the enzyme–substrate interaction shown as bonds (with carbon atoms in yellow). The modeled inhibitor is also shown as bonds (with carbon atoms in dark cyan). Hydrogen bonds are shown as green dotted lines. Note that the hydrogen bonds formed between the P1-Arg side-chain and the enzyme are apparently too short, indicating the necessity of an induced fit of the enzyme.

(Figure 5(a)). Asn192 is also located in a solvent-exposed loop region and, therefore, can be expected to undergo a moderate shift. As an example, residue 192 in coagulation factor Xa shifted about 2 Å as a result of the induced fit of the whole loop region during substrate binding.[31] However, the energy required for the induced fit would make the bulky P1 residue less preferable than small hydrophobic residues and if the energy cannot be balanced, such an accommodation is unlikely to happen.

Another energy-demanding event in the substrate binding would be the accommodation of a P2 residue other than Gly. As shown in the binding model, the side-chain of Tyr99 is in close contact with the side-chain of the P2-Pro, indicating that during actual substrate binding the side-chain of Tyr99 has to move away to make room for a P2 residue with long and/or bulky side-chain, as observed in kallikrein.[32] This is feasible in EFEa since the side-chain of Tyr99 is free and can swing to the spacious S4 subsite if the latter was not fully occupied by a large P4 residue. In fact, the superposition of the three independent EFEa molecules shows that the side-chain of Tyr99 is relatively mobile, with the O^{η} atoms of the phenolic side-chains deviating from each other by up to 1.65 Å (molecule A and C). The degree of the shift, and hence the energy required for the higher energy conformation, will vary for different P2 residues.

Correlation between structural and biochemical data

The model can be readily used to explain many of the confusing results from the substrate cleavage and inhibition experiments using both EFEa and F-II, a fibrinolytic enzyme component extremely homologous to EFEa (see Introduction), which provide strong evidence for our proposal. In the cleavage of bovine plasminogen (Li & He, unpublished results), EFEa cleaved the plasminogen at five inter-domain loops. Although a preference for small hydrophobic P1 residues and small P2 residues (Gly and Pro) was observed, the combination of P1-Arg and P2-Glu in one of the cleavage sites implied that enhanced binding of the loop with EFEa should have occurred during hydrolysis. In the substrate specificity assay, F-II failed to cleave the elastase-specific substrate Pyro-Glu-Pro-Val-pNA with a P2-Pro, but reacted with the thrombin-specific substrate Pyro-Glu-Gly-Arg-pNA with a P2-Gly, even though the latter substrate carried a P1-Arg less preferable than the P1-Val in the former substrate,[7] indicating the effect of S2 subsite occlusion. On the other hand, the observation that F-II was most reactive to substrate Meo-Suc-Ala-Ala-Pro-Val-pNA suggests that the increased binding energy in the enhanced binding of this prolonged substrate with the enzyme compensated for the energy required to shift Tyr99 away from its original position to make room for P2-Pro. However, even with this substrate, the amidolytic activity was still extremely low, in striking contrast to the high efficiency cleavage of macromolecules. The cleavage at even the non-specific sites (the peptide bonds behind Ser, Gly, and His) of the oxidized insulin B chain and β-amyloid 1–40 by F-II[8] suggested that maximum binding of these two polypeptides without significant secondary structures with F-II at the P subsites (and probably other regions as well) resulted in the sacrifice of the P1 selectivity. Detailed examination of the P2 and P4 residues of all the cleavage sites revealed that in most cases, a large P2 residue was accompanied by a small P4 residue, in agreement with the structural fact that a shift of the S2-Tyr99

side-chain towards the S4 pocket in response to a large P2 residue will reduce the size of the S4 pocket, making it suitable for accommodating only a smaller P4 residue.

In contrast to the complete inhibition of F-II by DFP, F-II is very weakly inhibited by ε-ACA and TLCK (with 90% and 80% residual activity, respectively) and not inhibited by TPCK,[7] indicating that the S1 pocket cannot accept a basic or aromatic P1 residue without induced fit triggered by the enhanced substrate binding, as explained by the EFEa structure. On the other hand, the extremely weak inhibition of F-II by elastatinal, a potent elastase inhibitor, might be due to the occlusion of Tyr99 to its bulky P2 group. The much stronger inhibition of F-II by the macromolecular trypsin-specific inhibitors SBTI and aprotinin (with 17% and 25% residual activity, respectively) implies that enhanced binding and induced fit of F-II should have occurred.

HLE has been reported to degrade fibrin initially at the C terminus of the Aα chain, then at the N terminus of the Aα chain, and then at the Bβ and γ chains.[33] Based on EFEa's structural similarity with HLE at the active site, we assumed that the characteristic incomplete fibrinolysis of EFEa observed in the fibrin plate assays could be explained, at least in part, by its elastase-like activity.

Biological implication

Our structural study revealed that EFEa possesses essential structural determinants characteristic of elastase and would, therefore, prefer elastase-specific small hydrophobic P1 residues. We have also demonstrated that the four-residue insertion after Ser217 and the extended S1 specificity pocket are crucial structural determinants responsible for its relatively broad substrate specificity. Our structure-based substrate modeling further suggested that EFEa would also be able to react with substrates carrying a basic or aromatic P1 residue if enhanced substrate binding and induced fit can be achieved. All these structural features and proposed mechanisms could explain the substrate specificity of EFEa, including the dual fibrinolysis.

Although the oral preparation containing all EFE components is clinically effective in thrombolysis,[5] EFEa's contribution to the therapy still remains unknown. As an oral-administered enzyme, EFEa must overcome many barriers, such as the intestinal wall,[34] before functioning in the blood stream. Nevertheless, the applicability of EFEa in the blood stream, a finely regulated system, appears to be a basic yet critical determinant regarding EFEa's potential for becoming a novel thrombolytic agent. Although the dual fibrinolysis of EFEa seems advantageous, the structural determinants featuring its elastase nature and relative broad substrate specificity have made pharmacological, biochemical and further structural studies of EFEa highly desirable and imperative.

EFE activity was detected in various parts of the earthworm digestive tract,[6] indicating that EFE components are digestive enzymes. From an evolutionary point of view, the broad substrate specificity of EFE components, as demonstrated by the currently studied EFEa and others,[8] is apparently very advantageous for earthworms to digest proteins efficiently under their specific living and nutritional conditions. Being the first crystal structure of an earthworm fibrinolytic enzyme component and serine protease originating from annelid worms, the EFEa structure reported here provides a template for accurately modeling other homologous family members.

Materials and Methods

EFEa purification, crystallization and heavy atom derivative preparation

The purification and crystallization of EFE component A was carried out as described earlier.[9] The crystallization was carried out at 15 °C using the hanging-drop vapor diffusion method. The drops were composed of 2.5 μl protein solution (10 mg/ml in d.d. water) and 2.5 μl mixing solution (1.2 M $(NH_4)_2SO_4$, 5.0% (v/v) PEG400, and 0.10 M Mops buffer (pH 7.2)) and were in equilibrium with 1 ml reservoir solution containing 2.0 M $(NH_4)_2SO_4$, 5.0% PEG400, and 0.10 M Mops buffer (pH 7.2). Good quality single crystals were obtained in ten days. Isomorphous heavy-atom derivative crystals were prepared by soaking the native crystals in the mother liquor containing 1.7 M Li_2SO_4 and 0.05 M Mops (pH 7.2) with the heavy-atom reagents dissolved at the specified concentration (Table 1).

Data collection and processing

X-ray diffraction data of the native and all heavy atom derivatives were collected at 18 °C using a Mar image plate system (MAR Research, Hamburg, Germany) with a sealed-tube X-ray generator operating at 40 kV and 50 mA. The reflection intensities were integrated and scaled by DENZO and SCALEPACK, respectively.[35]

The crystal belongs to the orthorhombic space group $P2_12_12_1$, with cell constants of $a = 40.6$ Å, $b = 126.1$ Å, $c = 129.1$ Å ($\alpha = \beta = \gamma = 90°$). According to the solvent content estimate, the crystal contains three molecules per asymmetric unit with a v/v solvent content of 45.0% ($V_m = 2.24$ Å^3 Da^{-1}).[36] The three molecules in the asymmetric unit were found to be related by two non-crystallographic 2-fold axes, the axis relating molecule A and B being substantially proper[37] and the axis relating molecule A and C being improper. These two axes are substantially parallel to *bc* plane and form an angle of about 30° with each other as viewed along the crystallographic axis *a*. The merged native data set (9.0 Å to 2.3 Å) consists of 29,703 unique reflections ($-3.0 \times \sigma$ cut-off), with an overall R_{merge} (on intensity) of 16.2% and completeness of 99.7% (98.9% complete between 2.35 Å and 2.3 Å). The diffraction data statistics for the native and heavy atom derivative crystals used in the MIR phasing are summarized in Table 1.

Table 1. Data collection and heavy-atom refinement statistics

Compound	Concentration (mM)	Time	Resolution (Å)[a]	No. of reflection	Completeness (%)	R_{sym}[b]	R_{deriv}[c]	No. of sites	Phasing power[d]	R_{cullis}[e]
Native	–	3 h[f]	2.3/2.5	23546	74/41	9.5	–	–	–	–
K_2PtCl_6	4.12	3 d	2.6/3.5	16930	74/42	10.2	15.2	3	1.39	0.59
$KAuCl_4$	1.30	3 d	2.7/3.5	16012	74/45	12.0	14.5	3	1.18	0.64
C_2H_5HgCl	2.00	3 d	2.4/2.5	20500	74/45	10.3	15.0	6	1.05	0.70
C_3H_7HgCl	1.81	3 d	3.0/3.0	10630	78/44	10.9	14.0	4	1.11	0.69

Reflections with $I > -3.0 \times \sigma_I$ were all used in the refinement, while a $2 \times \sigma_I$ cutoff was applied to generate the statistics in columns 5–8. The figure of merit calculated for all the 24,871 reflections within the resolution range 9.0–2.5 Å is 0.45.

[a] The resolutions behind "/" refer to the limits used in MIR phasing so they are applicable to both the phasing power and R_{cullis}.

[b] $R_{sym} = \sum|I - \langle I\rangle|/\sum\langle I\rangle$.

[c] $R_{deriv} = \sum(|F_{ph} - F_p|)/\sum F_p$.

[d] Phasing power $= [\sum|F_h|^2/\sum(|F_{ph}(obs)| - |F_{ph(calc)}|)^2]^{1/2}$, for centric reflections.

[e] $R_{cullis} = \sum||F_{ph} \pm F_p| - F_{ph(calc)}|/\sum|F_{ph} \pm F_p|$.

[f] Native crystal was also soaked in the mother liquor without heavy-atom regent for three hours.

MIR phasing

The EFEa structure was solved using the MIR method. The heavy atom positions in the K_2PtCl_6 derivative were determined in the difference Patterson map calculated with 9.0–3.0 Å data using the CCP4 program package.[38] The heavy atom positions in the other derivatives were located in the difference Fourier map calculated using the K_2PtCl_6-derived phases. The heavy-atom parameters were refined and the MIR phases were calculated to 2.5 Å resolution with the MLPHARE program[38] (Table 1).

Density modification, map calculation, and NCS averaging

The initial phases were improved by solvent flattening and histogram matching with the DM program in the CCP4 package.[38] NCS matrices determined using the heavy atom sites were used to generate a molecular envelope around molecule A using MAMA.[39] The matrices were further improved by IMP[39] before NCS averaging was performed using SPANCSI.[39] Solvent-flattened and NCS-averaged maps at various resolutions were generated and used in the model building.

Model building and refinement

An *ab initio* model was built with the O[40] program on a SGI workstation. The EFEa molecule backbone was traced without much difficulty. The sequence of the earthworm fibrinolytic enzyme from *L. bimastus*, an enzyme extremely homologous to EFEa (see Introduction) was used to generate the initial complete molecule with residue 219 corrected after the EFEa sequence was determined. The model was refined using molecular dynamics at 3.0 Å. All refinements were performed with the program X-PLOR[41] using the stereochemical parameters of Engh & Huber[42] and $0 \times \sigma_F$ cutoff. Difference or omit electron density maps were calculated and inspected after each refinement and the model was adjusted manually when necessary. Positional refinement was continued as the resolution was gradually extended from 3.0 to 2.3 Å. At 2.3 Å, a NCS-restrained simulated annealing refinement was performed using different weighting criteria for different segments of the polypeptide according to the judgment of mobility. The NCS restraint was then released. After the *R* factor was reduced to below 24.0%, some peaks above at least $3.0 \times \sigma$ in the $F_o - F_c$ electron density map were selected as candidates for water molecules. The individual *B* factors were refined using 2.3 Å data.

Model quality and accuracy

The refined model gives an *R* factor of 19.1% for 9.0–2.3 Å data (27,322 reflections with $F > 0 \times \sigma_F$) and a free *R* factor of 23.6% (2382 reflections). It includes all the 723 residues of the three EFEa molecules and 278 water molecules in the asymmetric unit. Structural evaluation with the PROCHECK[43] program indicates that the refined structure has good geometric parameters (Table 2), with 88% of the residues lying in the most favored region and the rest in the allowed region of the Ramachandran plot. An example of the final $2F_o - F_c$ electron-density map calculated with the model phases is given in Figure 4. Almost all residues are well defined in the final $2F_o - F_c$ map. Some of the side-chain atoms of Gln36, Leu60C, Asn93 and Leu217D are undefined

Table 2. Crystallographic refinement statistics

Data resolution (Å)	9.0–2.3
Number of reflections with $F > 0 \times \sigma_F$ (all)	29,703
Number of reflections with $F > 0 \times \sigma_F$ (test)	2382
R factor (%)[a]	19.1
Free *R* factor (%)[b]	23.6
Non-hydrogen protein atoms[c]	1729 × 3
Number of water molecules	278
RMS standard deviations	
Bond length (Å)	0.011
Bond angles (°)	1.63
Dihedral angles (°)	28.14
Improper angles (°)	0.87
B factors for bonded atoms (Å^2)	2.6
Average *B* factors (Å^2)	
Main chain (molecule A/B/C)	21.8/16.5/17.5
Side-chain (molecule A/B/C)	21.8/16.4/17.4
Water molecules	30.3

[a] *R* factor $= (\sum||F_{obs}| - |F_{calc}||/\sum|F_{obs}|) \times 100$, where F_{obs} and F_{calc} are the observed and calculated structure factor amplitudes, respectively.

[b] A small fraction (8%) of reflections were randomly selected and used to calculate the free *R* factor.

[c] There are three EFEa molecules in the asymmetric unit with each molecule containing 1729 non-hydrogen protein atoms.

and, therefore, excluded from the phasing, but they are still included in the coordinate set.

The secondary structure of EFEa (Figure 1) was calculated with STRIDE.[44] Figure 1 was produced with Alscript,[45] Figures 2, 3(c), 4, 5(b) and 6 with Molscript[46] and Raster3d,[47]Figure 3(a) and (b) with GRASP[48] and Figure 5 with Turbo-Frodo.[49]

Accession numbers

The coordinates and structure factor amplitudes for EFEa have been deposited in the Protein Data Bank with accession code 1IJ7.

Acknowledgments

This project was supported by the Key Project Foundation of the Chinese Academy of Sciences (Project No. KJ951-A1-601), the National Natural Sciences Foundation of China (Project No. B705975), and the National Key Research Development Project of China (Project No. G1999075601).

References

1. Lijnen, H. R. & Collen, D. (1995). Fibrinolytic agents: mechanisms of activity and pharmacology. *Thromb. Haemostat.* **74**, 387–390.
2. Mihara, H., Sumi, H., Akazawa, K., Yoneds, T. & Mizunoto, H. (1983). Fibrinolytic enzyme extracted from the earthworm. *Thromb. Haemostat.* **50**, 258.
3. Zhou, Y. C., Zhou, H. & Chen, Y. C. (1988). Purification and biochemical characterization of the fibrinolytic enzymes from the earthworm *Eisenia fetida*. *Acta Biochem. Biophys. Sinica*, **20**, 35–42.
4. Daedo Pamphlet (1990). YONGSHIM capsule, pp. 1–2, Daedo Pharmaceutical Co., Ltd., Korea.
5. Hou, Q. (1995). Mass production of earthworm fibrinolytic enzyme under GMP standard. *Prog. Biochem. Biophys. (Beijing)*, **22**, 567.
6. Mihara, H., Sumi, H., Yoneta, T., Mizumoto, H., Ikeda, R., Seiki, M. & Maruyama, M. (1991). A novel fibrinolytic enzyme extracted from the earthworm *Lumbricus rubellus*. *Jpn J. Physiol.* **41**, 461–472.
7. Nakajima, N., Mihara, H. & Sumi, H. (1993). Characterization of potent fibrinolytic enzymes in earthworm, *Lumbricus rubellus*. *Biosci. Biotech. Biochem.* **57**, 1726–1730.
8. Nakajima, N., Sugimoto, M., Ishihara, K., Nakamura, K. & Hamada, H. (1999). Further characterization of earthworm serine proteases: cleavage specificity against peptide substrates and on autolysis. *Biosci. Biotechnol. Biochem.* **63**, 2031–2033.
9. Tang, Y., Zhang, J., Gui, L., Wu, C., Fan, R., Chang, W. & Liang, D. (2000). Crystallization and preliminary X-ray analysis of earthworm fibrinolytic enzyme component A from *Eisenia fetida*. *Acta Crystallog. sect. D*, **56**, 1659–1661.
10. Schechter, I. & Berger, A. (1967). On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **27**, 157–162.
11. Hartley, B. S. & Shotton, D. M. (1971). Pancreatic elastase. In *The Enzymes* (Boyer, P. D., ed.), vol. 3, pp. 323–373, Academic Press, New York, NY.
12. Takahashi, L. H., Radhakrishnan, R., Rosenfield, R. E., Meyer, E. F. & Trainor, D. A. (1989). Crystal structure of the covalent complex formed by a peptidyl alpha,alpha-difluoro-beta-keto amide with porcine pancreatic elastase at 1.78-angstroms resolution. *J. Am. Chem. Soc.* **111**, 3368–3374.
13. Wei, A.-Z., Mayr, I. & Bode, W. (1988). The refined 2.3 angstroms crystal structure of human leukocyte elastase in a complex with a valine chloromythyl ketone inhibitor. *FEBS Letters*, **234**, 367–373.
14. Marquart, M., Walter, J., Deisenhofer, J., Bode, W. & Huber, R. (1983). The geometry of the reactive site and of the peptide groups in trypsin, trypsinogen and its complexes with inhibitors. *Acta Crystallog. sect. B*, **39**, 480–490.
15. Mac Sweeney, A., Birrane, G., Walsh, M. A., O'Connell, T., Malthouse, J. P. & Higgins, T. M. (2000). Crystal structure of delta-chymotrypsin bound to a peptidyl chloromethyl ketone inhibitor. *Acta Crystallog. sect. D*, **56**, 280–286.
16. Spraggon, G., Phillips, C., Nowak, U. K., Ponting, C. P., Saunders, D., Dobson, C. M. *et al.* (1995). The crystal structure of the catalytic domain of human urokinase-type plasminogen activator. *Structure*, **3**, 681–691.
17. Lamba, D., Bauer, M., Huber, R., Fischer, S., Rudolph, R., Kohnert, U. & Bode, W. (1996). The 2.3 Å crystal structure of the catalytic domain of recombinant two-chain human tissue-type plasminogen activator. *J. Mol. Biol.* **258**, 117–135.
18. Wang, X. Q., Lin, X. L., Loy, J. A., Tang, J. & Zhang, X. J. (1998). Crystal structure of the catalytic domain of human plasmin complexed with streptokinase. *Science*, **281**, 1662–1665.
19. Sugimoto, M. & Nakajima, N. (2001). Molecular cloning, sequencing, and characterization of cDNAs encoding fibrinolytic enzymes from earthworm, *Lumbricus rubellus*. *Biosci. Biotechnol. Biochem.* **65**, 1575–1580.
20. Robertus, J., Kraut, J., Alden, R. A. & Birktoft, J. J. (1972). Subtilisin: a stereochemical mechanism involving transition-state stabilization. *Biochemistry*, **11**, 4293–4303.
21. Wright, H. T. (1973). Activation of chymotrypsinogen-A: an hypothesis based upon the comparison of the crystal structures of chymptrypsinogen-A and α-chymotrypsin. *J. Mol. Biol.* **79**, 13–23.
22. Huber, R. & Bode, W. (1978). Structural basis of the activation and action of trypsin. *Accts Chem. Res.* **11**, 114–122.
23. Madison, E. L., Kobe, A., Gething, M. J., Sambrok, J. F. & Goldsmith, E. J. (1993). Converting tissue plasminogen activator to a zymogen: a regulatory triad of Asp-His-Ser. *Science*, **262**, 419–421.
24. Bode, W., Meyer, E. & Power, J. C. (1989). Human leukocyte and porcine pancreatic elastase: X-ray crystal structures, mechanism, substrate specificity, and mechanism-based inhibitors. *Biochemistry*, **28**, 1951–1963.
25. Greer, J. (1990). Comparative model-building of the mammalian serine proteases. *Proteins*, **7**, 317–334.
26. Perona, J. J. & Craik, C. S. (1995). Structural basis of substrate specificity in the serine proteases. *Protein Sci.* **4**, 337–360.
27. Madison, E. L., Goldsmith, E. J., Gerard, R. D., Gething, M.-J. H. & Sambrook, J. F. (1989). Serpin-resistant mutants of human tissue-type plasminogen activator. *Nature*, **339**, 721–724.

28. Madison, E. L., Goldsmith, E. J., Gerard, R. D., Gething, M.-J. H., Sambrook, J. F. & Bassel-Duby, R. S. (1990). Amino acid residues that affect interaction of tissue-type plasminogen activator with plasminogen activator 1. *Proc. Natl Acad. Sci. USA*, **87**, 3530–3533.
29. Bennett, W. F., Paoni, N. F., Keyt, B. A., Botstein, D., Jones, A. J., Presta, L. *et al.* (1991). High resolution analysis of functional determinant determinants on human tissue-type plasminogen activator. *J. Biol. Chem.* **266**, 5191–5201.
30. Bode, W. & Huber, R. (1992). Natural protein proteinase inhibitors and their interaction with proteinases. *Eur. J. Biochem.* **204**, 433–451.
31. Brandstetter, H., Kuhne, A., Bode, W., Huber, R., von der Saal, W., Wirthensohn, K. & Engh, R. A. (1996). X-ray structure of active site-inhibited clotting factor Xa: implications for drug design and substrate recognition. *J. Biol. Chem.* **271**, 29988–29992.
32. Chen, Z. & Bode, W. (1983). Refined 2.5 Å X-ray crystal structure of the complex formed by porcine kallikrein and the bovine pancreatic trypsin inhibitor. *J. Mol. Biol.* **164**, 283–311.
33. Bach-Gansmo, E. T., Halvorsen, S., Godal, H. C. & Skjonsberg, O. H. (1994). Degradation of the alpha-chain of fibrin by human neutrophil elastase reduces the stimulating effect of fibrin on plasminogen activation. *Thromb. Res.* **75**, 307–317.
34. Fan, Q., Wu, C., Li, L., Fan, R., Wu, C., Hou, Q. & He, R. (2001). Some features of intestinal absorption of intact fibrinolytic enzyme III-1 from *Lumbricus rubellus*. *Biochim. Biophys. Acta*, **1526**, 286–292.
35. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326.
36. Matthews, B. W. (1968). Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491–497.
37. Vellieux, F. M. D. & Read, R. J. (1997). Noncrystallographic symmetry averaging in phase refinement and extension. *Methods Enzymol.* **277**, 18–53.
38. CCP4 (1994). CCP4 suite: programs for protein crystallography. *Acta Crystallog. sect. D*, **50**, 760–763.
39. Jones, T. A. (1992). A, yaap,asap, @#*? A set of averaging programs. In *Molecular Replacement* (CCP4), pp. 92–105, SERC Daresbury Laboratory, Warrington, UK.
40. Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and location of errors in these models. *Acta Crystallog. sect. A*, **47**, 110–119.
41. Brünger, A. T. (1996). *X-PLOR (Version 3.851). A System for X-ray Crystallography and NMR*, Yale University Press, New Haven, CT.
42. Engh, R. A. & Huber, R. (1991). Accurate bond and angle parameters for X-ray protein structure and refinement. *Acta Crystallog. sect. A*, **47**, 392–400.
43. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283–291.
44. Frishman, D. & Argos, P. (1995). Knowledge-based secondary structure assignment. *Proteins: Struct. Funct. Genet.* **23**, 566–579.
45. Barton, G. J. (1993). ALSCRIPT: a tool to format multiple sequence alignments. *Protein Eng.* **6**, 37–40.
46. Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946–950.
47. Merritt, E. A. & Murphy, M. E. P. (1994). Raster3D version 2.0. A program for photorealistic molecular graphics. *Acta Crystallog. sect. D*, **50**, 869–873.
48. Nicholls, A., Bharadwai, R. & Honig, B. (1993). GRASP—graphical representation and analysis of surface properties. *Biophys. J.* **64**, A166.
49. Roussel, A. & Cambillau, C. (1991). "Turbo Frodo". In *Silicon Graphics Geometry Partners Directory*, p. 86, Silicon Graphics, Mountain View, CA.
50. Renatus, M., Engh, R. A., Stubbs, M. T., Huber, R., Fischer, S., Kohnert, U. & Bode, W. (1997). Lysince 156 promotes the anomalous proenzyme activity of tPA: X-ray crystal structure of single-chain human tPA. *EMBO J.* **16**, 4797–4805.

Edited by R. Huber

(Received 21 November 2001; received in revised form 21 May 2002; accepted 28 May 2002)