Available online at www.sciencedirect.com
SCIENCE DIRECT®
Biochemical and Biophysical Research Communications 325 (2004) 877–882
BBRC
www.elsevier.com/locate/ybbrc

Structural basis for broad substrate specificity of earthworm fibrinolytic enzyme component A

Chao Wang[a,b], Feng Wang[a,b], Mei Li[a], Yong Tang[a], Ji-Ping Zhang[a], Lu-Lu Gui[a], Xiao-Min An[a], Wen-Rui Chang[a,*]

[a] National Key Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, 15 Datun Road, Chaoyang District, Beijing 100101, PR China
[b] Graduate School of the Chinese Academy of Sciences, Beijing 100101, PR China

Received 13 October 2004
Available online 5 November 2004

Abstract

Earthworm fibrinolytic enzyme component A (EFE-a) possesses an S1[1] pocket, which is typical for an elastase-like enzyme, but it can still hydrolyze varieties of substrates, and it exhibits wide substrate specificity. Former structure studies suggested that the four-residue insertion after Val[217] [2] might endow EFE-a with this specificity. Based on the native crystal structure at a resolution of 2.3 Å, we improved the native crystal structure to 1.8 Å and determined its complex structure with the inhibitor Meo-Suc-Ala-Ala-Pro-Val-CMK at a resolution of 1.9 Å. The final structures show that: (1) EFE-a possesses multisubstrate-binding sites interacting with the substrates; (2) significant conformation adjustment takes place at two loops binding to the N-terminal of the substrates, which may enhance the interaction between the enzyme and the substrates. These characteristics make the substrate-specificity of EFE-a less dependent on the property of its S1-pocket and may endow the enzyme with the ability to hydrolyze chymotrypsin-specific substrates and even trypsin-specific substrates.

Keywords: Earthworm fibrinolytic enzyme; Substrate specificity; Multisubstrate-binding sites

Earthworm fibrinolytic enzyme (EFE), which is cheap and easily stored, has been widely used as a novel, orally administered drug for thrombosis treatment and has aroused great interest in East Asia. Research has shown that EFE contains several active components [1–4]. We systematically isolated and purified seven different components from earthworm *Eisenia fetida*, and characterized them [5]; then we determined the structure of one of its components—earthworm fibrinolytic enzyme component A (EFE-a)—at a resolution of 2.3 Å (PBD: 1M9U). This was the first crystal structure from earthworm. Based on its crystal structure, we discussed the reason that EFE-a had dual fibrinolytic activity [6,7].

EFE-a is a single-peptide-chained serine protease composed of 241 amino acid residues and it belongs to the S1A subfamily. It has a molecular weight of 24,663 Da (MALDI-TOF result) and an isoelectric point of 3.5. Its highly hydrophobic S1-pocket, which has the primary specificity determinants of Gly[189], Val[216], and Thr[226], is preferable for elastase-specific small hydrophobic P1 residues, which suggests that EFE-a should be a typical elastase [6]. However, chromogenic substrate hydrolyzing studies showed that EFE-a also cleaved chymotrypsin-specific and

* Corresponding author. Fax: +86 10 64889867.
E-mail address: wrchang@sun5.ibp.ac.cn (W.-R. Chang).

[1] Nomenclature for the substrate amino acid residues is P*n*, ..., P2, P1, P1′, P2′, ..., P*n*′, where P1–P1′ denotes the hydrolyzed bond. S*n*, ..., S2, S1, S1′, S2′, ..., S*n*′ denotes the corresponding enzyme binding sites.

[2] Chymotrypsinogen numbering is used throughout.

0006-291X/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.bbrc.2004.10.113

trypsin-specific substrates in addition to elastase-specific ones [5]. On the basis of its native crystal structure at the resolution of 2.3 Å, we built an eight-residue substrate $(Ala)_6$-Pro-Arg using the program O [8] and simulated its interaction with EFE-a. The result indicated that the four-residue insertion after Val^{217}, which extends the β strand 11, enlarges the S1-pocket, and provides three more pairs of hydrogen bonds between EFE-a and the substrates, may endow EFE-a with the ability to hydrolyze chymotrypsin-specific and trypsin-specific substrates [6]. In this paper, we report the native crystal structure of EFE-a at a resolution of 1.8 Å and its complex crystal structure with an inhibitor at a resolution of 1.9 Å, and discuss the structural basis for its wide substrate specificity.

Materials and methods

Purification and crystallization. The purification [5] and crystallization [7] of EFE-a were carried out as described previously. The crystal was grown at 288 K by hanging-drop vapor-diffusion. The drops contained a mixture of 2.5 μl protein solution (10 mg/ml) with 2.5 μl precipitant solution [1.2 mol/L $(NH_4)_2SO_4$, 5.0% (v/v) PEG400, and 0.10 mol/L Mops (pH 7.2)], and the reservoir consisted of 1.0 ml of 2.0 mol/L $(NH_4)_2SO_4$, 5.0% (v/v) PEG400, and 0.10 mol/L Mops (pH7.2). Crystals suitable for diffraction could be obtained in 10 days. Complex crystals were prepared by soaking native crystals in the mother liquid (1.6 mol/L $(NH_4)_2SO_4$, 5.0% (v/v) PEG400, and 0.10 mol/L Mops (pH 7.2) containing 4 mmol/L inhibitor MeO-Suc-Ala-Ala-Pro-Val-CMK for 2 days. To increase the solubility of the inhibitor, some ethanol was added to the mother liquid and the final concentration of the ethanol was 2% (v/v).

Data collecting and processing. Diffraction data were collected at the synchrotron of the Photon Factory, Tsukuba, Japan. The native data were collected at room temperature 291 K, and the complex data were collected at 100 K. The data were processed using the programs *DENZO* and *SCALEPACK*. Details are shown in Table 1.

Structure determination and refinement. Taking the 2.3 Å native crystal structure of EFE-a (PDB: 1M9U) as an initial model, we extended the crystal resolution to 1.8 Å. After rigid body refinement, the *R* factor was 29.7% and the *R*-free was 30.0%. The structure was finally refined to an *R* factor of 17.6% and an *R*-free of 20.3%.

The complex structure was solved by the program MolRep [9] using the native crystal structure as a search model, since the unit cell parameters of the complex crystal had changed greatly compared with those of the native crystal. After rigid body and anneal refinement, the *R* factor of the structure was 30.1% and the *R*-free was 33.4%. We built the model of the inhibitor referring to the complex crystal structure of human leukocyte elastase with its inhibitor (PDB: 1PPG), and obtained the topology and parameter files using the program Xplor2d. Since the inhibitor covalently bound to the enzyme, the covalent bonds were defined referring to the definition of disulfide bonds in the CNS program. After one round of manual fitting in the O program, the inhibitor was introduced into the obtained model based on a high quality density map. Another round of anneal refinement was carried out and resulted in an *R* factor of 26.7% and an R-free of 29.4%. The structure was finally refined to an *R* factor of 19.4% and an *R*-free of 21.7%. All model rebuilding and adjustments were carried out using the program *O* [8] and all the model refinements and map calculations were performed using the program *CNS*.

Stereochemistry of refined models was checked by the program *PROCHECK* [10] and no residues were found in the disallowed region of the Ramachandran plot in any of the models. The secondary structure was calculated by the program *STRIDE*. Table 1 shows a summary of the refinement statistics of all structures.

Table 1
Data collection and refinement statistics

	Native	Complex
Data collection		
Resolution (Å)	40–1.80	40–1.90
Space group	$P2_12_12_1$	$P2_12_12_1$
Cell constants (Å)	$a = 40.6$	$a = 40.2$
	$b = 127.5$	$b = 126.3$
	$c = 129.2$	$c = 127.3$
Number of molecules per asymmetric unit	3	3
Solvent content (%)	45.4	43.3
Total reflections	379,493	433,973
Unique reflections	59,211	50,097
Completeness (%)	93.2 (86.3)	95.9 (92.7)
I/σ	21.89 (3.71)	14.55 (2.95)
R-merge (%)	8.0 (48.5)	8.5 (43.8)
Refinement statistics		
R-work (%)	17.6	19.4
R-free (%)	20.3	21.6
Number of non-hydrogen atoms	5551	5860
Number of water	354	568
RMSD		
Bond length (Å)	0.004	0.007
Bond angle (°)	1.26	1.45
Average *B* factor (Å^2)		
Protein main chain	18.43	18.17
Protein side chain	18.83	18.53
Water	25.15	26.57
Ramachandran (%)		
Most favored	88.7	87.9
Additional favored	11.3	12.1

Values in parentheses show the last resolution shell in Å.

Results and discussion

Comparison of native crystal structures at two different resolutions

Both the native and complex crystal structures of EFE-a have three molecules in one asymmetric unit, named A, B, and C. Superimposing the Cα carbon atom of the counterpart molecules (A, B, and C) between the crystal structure at the resolution of 2.3 Å (1M9U) and the structure at 1.8Å using a 0.5 Å distance cutoff resulted in 233, 236, and 237 topologically equivalent atoms with root mean square (RMS) deviations of 0.18, 0.16, and 0.18 Å, respectively, which indicated again that the former structure was reliable. The structure at a higher resolution could give us more details so all the following discussions are based on the structure at 1.8 Å.

The side-chain of residue Asn^{62} of all three molecules in the native structures had poor density both in $2F_o - F_c$ and the Omit map, and the same thing happened in the complex structure. Based only on the

density map, we thought the residue should be a serine. However, due to the lack of further evidence from the cDNA sequence and MS [5], we remained this residue as an asparagine.

By computer simulation we found that the side-chain of Tyr^{99} was in close contact with the side-chain of the P2-Pro, indicating that during actual substrate binding, the side-chain of Tyr^{99} would have to move away to make room for a P2 residue with a long and/or bulky side-chain, which would be an energy-consuming process. Investigating the conformation of Tyr^{99} in all three molecules of the native structure we found that the phenolic side-chain of Tyr^{99} took different conformation in different molecules, and its $2F_o - F_c$ density was very dispersed. These findings imply that it has high flexibility and can swing away spontaneously to accommodate P2 residues with long and/or bulky side-chains into S2 site during the substrate binding.

Multisubstrate-binding sites and substrate specificity

The canonical "Key and Lock" model assumed that the accommodation ability of the S1-pocket was the main determinant of the substrate specificity of a serine protease [11]. EFE-a possesses a typical S1-pocket of an elastase-like protease, which suggests that it should be preferable only for P1 residues with small and hydrophobic side-chains. However, chromogenic substrate hydrolyzing studies showed that, compared to porcine pancreatic elastase (PPE), EFE-a had rather low cleavage activity toward an elastase-specific substrate, and its efficiency in cleaving a substrate with P1-Phe/Arg (specific for chymotrypsin/trypsin) was three to four times greater than its efficiency in cleaving a substrate with P1-Ala/Val (specific for elastase) [5]. These facts indicated that besides having some ability to hydrolyze elastase-specific substrates, EFE-a also has a relatively stronger ability to hydrolyze chymotrypsin-specific and trypsin-specific substrates, and shows broad substrate specificity. The results of hydrolyzing fibrin(ogen), plasminogen, and oxidized insulin β-chain also demonstrated the wide substrate specificity of EFE-a [4,12].

By superimposing the complex structure to the native one, we found that the four-residue inhibitor not only interacted with the enzyme at the S1-pocket by inserting P1-Val residue into the pocket, but also at the S2-, S3-, and S4-sites which, respectively, interacted with the P2, P3, and P4 residues of the inhibitor (Fig. 1).

We have discussed previously that the phenolic side-chain of Tyr^{99} was somewhat flexible. We assume that this flexibility endows the S2-site with more plasticity, which may make the S2-site able to accommodate residues with not only P2-Gly/Ala but also P2-Pro. Definite movement of the Tyr^{99} side-chain has been observed in the complex structure. And together with Try^{94} and His^{57} it formed a bowl-shaped S2-site interacting with the side-chain of P2-Pro by hydrophobic interaction.

P3-Ala of the inhibitor interacted with EFE-a at Val^{216}, which was called an S3-site, through two pairs of hydrogen bonds: between O of Ala^{P3} and N of Val^{216}, and between N of Ala^{P3} and O of Val^{216}. In PPE, Arg^{217A} (Arg^{217} in human leukocyte elastase, HLE) interacted with the P4-site of the substrate, while in EFE-a, the residue at 217 was Val, which could not interact with P4-Ala of the substrate. Instead, a hydrogen bond between N of Ala^{P4} and O of Gly^{173D} associated the inhibitor to the enzyme. Gly^{173C}, Gly^{173D}, Ser^{98B}, Tyr^{99}, and Trp^{215} together formed the S4-site of EFE-a.

Further optimal superposition of the EFE-a complex structure with the PPE complex structure (PDB: 1QR3)

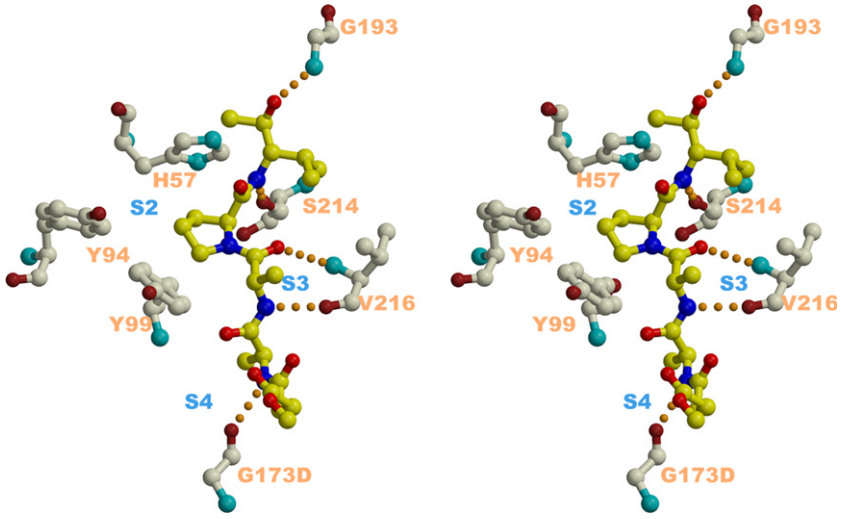

Fig. 1. The substrate-binding sites S2–S4 of EFE-a.

and HLE complex structure (PDB: 1PPG) showed us that two other substrate binding sites S1′ (Cys^{42}–Cys^{58} and Thr^{41}) and S2′ (His^{40}, Arg^{143}, and Leu^{151}) could also be found in EFE-a which, respectively, interacted with the P1′ and P2′ residues of a substrate.

The results of former computer simulating indicated that the side-chain of P1-Arg could not insert into the S1-pocket with a normal conformation; instead, it could "sit" upon the pocket in a "bent" manner [6], which was obviously energy-unfavorable. In this study, we think that the binding of the substrates to EFE-a might mainly depend not on the specificity of the S1-pocket but on the collaboration of those multisubstrate-binding sites. When the energy is unfavorable on one site, while the others could compensate and make the total energy favorable, then the substrates may bind to the enzyme. Recent studies on the kinetics of EFE-a found that it had a K_m for Tos-Gly-Pro-Arg-4-pNA (the chromogenic substrate for trypsin) of 60 μmol/L, while for Suc-Ala-Ala-Ala-pNA (the standard chromogenic substrate of elastase), it had a K_m of 246 μmol/L [12], which indicated that its typical elastase-specific S1-pocket did not play a predominant role in the process of substrates binding to EFE-a.

Surface loops and substrate specificity

Serine proteases in the S1A sub-family have many surface loops, some of which take part in recognition and binding of substrates. Previous engineering experiments investigating these enzymes have revealed that these surface loops are incredibly important in the determination of the substrate specificity of enzymes; some mutants, by simply substituting the S1-pocket's specificity determinant residues, could not change the substrate specificity of the enzymes, while some other surface-loop-hybrid mutants could [13,14]. As EFE-a possesses a typical elastase-specific S1-pocket while it does not have substrate specificity typical for elastase-like enzymes, we suggested that the surface loops of EFE-a may play some dominant role in the substrate specificity determination.

Compared with PPE and HLE, EFE-a has three quite different surface loops: His^{91}–Ala^{104} (named Loop C), Thr^{165}–His^{180} (named Loop 3), and Val^{217}–Ser^{226} (named Loop 2). Loop 2 and Loop 3 are much longer than the counterpart loops of PPE and HLE, and Loop C takes a different orientation from the counterpart loops in PPE and HLE (Fig. 2). Compared with bovine α-chymotrypsin, all these loops have some amino acid insertions: SSGL after Val^{217} in Loop 2, GVGG after Val^{173} in Loop 3, and AS after Thr^{98} in Loop C.

The four-residue insertion in Loop 2 and Loop 3 makes these two loops more flexible. By superimposing the complex structure of EFE-a with its native structure, we found that after the inhibitor binding, some significant conformation changes occurred both at Loop 2 and Loop 3. And an obvious movement of the residue Gly^{173D} (Fig. 3) was also observed; in the native structure, O of Gly^{173D} (which is in the middle of Loop 3) formed a hydrogen bond with OD1 of Asn^{175} and OG of Ser^{98B} to stabilize the conformation of Loop 3; while in the complex one, Gly^{173D} swung away to the residue of P4-Ala, the former two hydrogen bonds were broken away, and instead, O of Gly^{173D} formed a new hydrogen bond with N of Ala^{P4} to assist the substrate binding to EFE-a. The conformation adjustments of Loop 2 and Loop 3 in EFE-a indicate that the substrate binding is an interactive process; when the substrate is present, the conformational adjustment of Loop 2 and Loop 3 is induced to take place, making EFE-a more accessible to the substrate. These two loops may perform like an adjustable pair of "forceps" that could hold the P4 and P5 residues of the substrates and then assist the binding of the substrates.

One of the most obvious differences between EFE-a and PPE (or HLE) is Loop 3. In EFE-a, Loop 3 has a four-residue insertion after Val^{173} and an α-helix

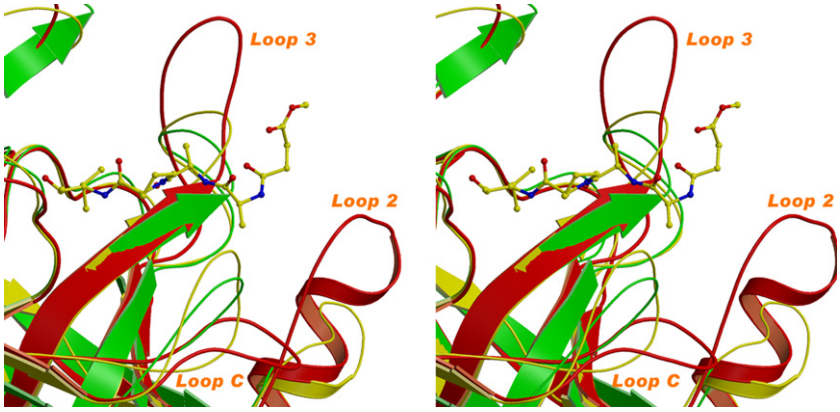

Fig. 2. The comparison of some loops of EFE-a with PPE and HLE loops. EFE-a is shown in red, PPE in yellow, and HLE in green. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

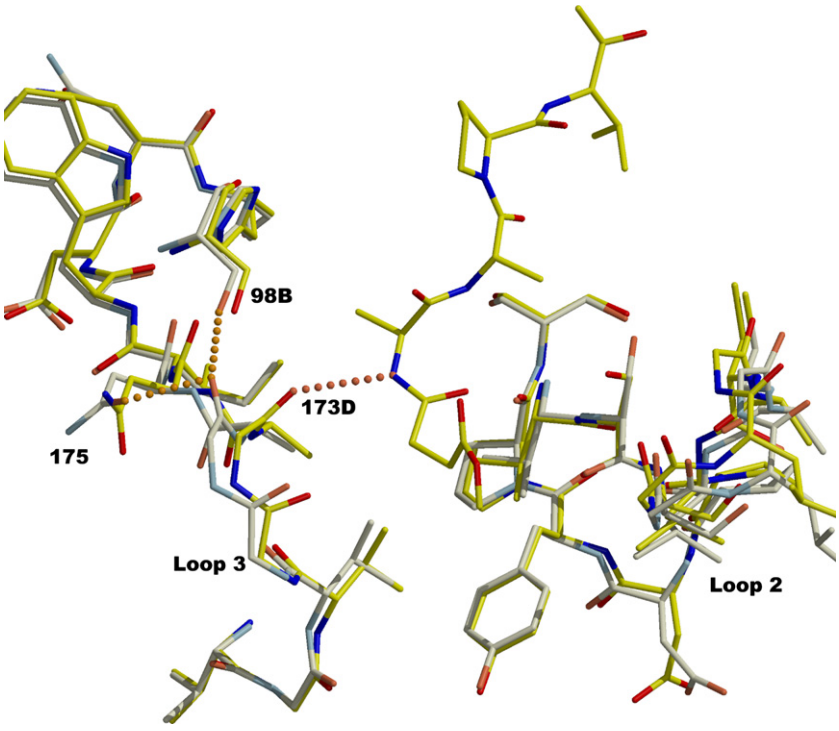

Fig. 3. The conformation changes of Loop B and Loop C in substrate binding. Superposing of crystal structures of EFE-a and its complex shows that in the substrate-binding course conformational adjustments of Loop B and Loop C will take place, which will facilitate the substrate binding to the enzyme.

terminated at Gly^{173A}; in PPE, this counterpart loop has 2 two-residue insertion after Ser^{170} and also an α-helix terminated at Ser^{170}; while in HLE, the residues from 169 to 176 are missing and there is no α-helix in this counterpart loop. By comparison we found that Loop 3 of EFE-a protrudes out of the enzyme surface. If the substrates tried to enter the catalytic site of EFE-a in a naturally extended manner, there would be obvious steric hindrance between the loop and the P4/P5 residues. In EFE-a's complex structure, the inhibitor did not bind to the catalytic site in a naturally extended manner, but bent at P4-Ala, following Loop 2, and the inhibitor was inserted into the valley formed by Loop 2 and Loop 3 (Fig. 4). In contrast, these events did not happen in PPE or HLE. As the substrates had to enter the catalytic site in this special way, we think that some conformational adjustments of a substrate have to take place in order to fit the shape of the catalytic site during its binding. And this "adjust-to-fit" model might partly result in the low hydrolytic efficiency of EFE-a.

The Loop C of EFE-a is also different from the counterpart loop of PPE or HLE. In all serine proteases of the S1A sub-family, Loop C and Loop 2 are separated into two different domains: the S2-site and S3-site are, respectively, located at Loop C and Loop 2, and both of the loops are close to the S1-pocket, which is important for the substrate recognition and locating scissile bonds. In PPE or HLE, this loop is oriented to the catalytic site and to the Loop 2, while in EFE-a, Loop C is far away from the catalytic site, and the substrate may be pushed to the Loop 2 side by the big side-chain of Tyr^{99}. It is Loop 3, instead of Loop C, which together with Loop 2 enables some interaction with the substrate. We assume that the lack of collaboration of Loop C and Loop 2 for substrate binding may be responsible for EFE-a having low hydrolyzing activity but wide substrate specificity.

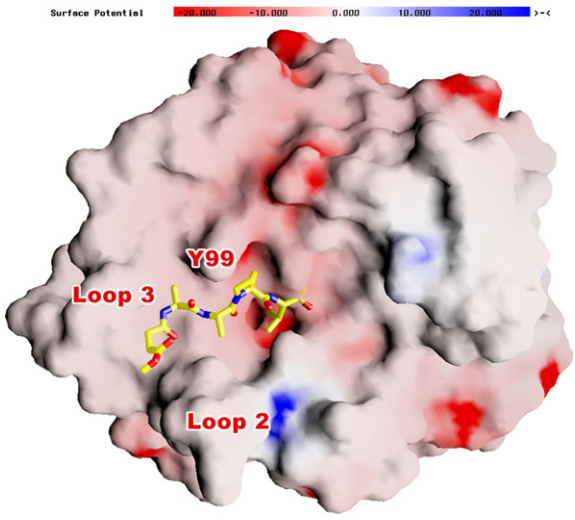

Fig. 4. The N-terminal substrate-binding valley of EFE-a. The N-terminal substrate-binding valley is composed of residue Y99, Loop 2, and Loop 3. Due to the steric hindrance of Loop 3, residues P4 and P5 of the substrate or inhibitor cannot access the valley in a natural manner, but bend at P4-Ala following Loop 2.

In addition to this inhibitor complex, we also tried to grow more complex crystals with other three-residue inhibitors containing P1-Lys/Arg, but have not succeeded yet. We assume that for short-chained substrates, the substrate specificity of EFE-a might be dominated by the specificity of the S1-pocket, which is well known as the canonical ''Key and Lock'' model, while for long-chained substrates, its substrate specificity may be dominated by the multisubstrate-binding sites and special surface loops.

Fibrin hydrolyzing test in vitro showed that EFE might take the same fibrinolytic strategy as HLE and could cut plasminogen into plasmin or mini-plasmin [12], which implies the possibility of making EFE-a, a novel drug for thrombosis treatment, although its wide substrate specificity might be the biggest hindrance to making this application a reality.

Acknowledgments

We are grateful to Prof. N. Sakabe and Dr. K. Sakabe of the Photon Factory (Tsukuba, Japan) for helping in data collection. This project was supported by the National Natural Sciences Foundation (No. 39970174), the National Key Research Development Project of China (No. G1999075601) and the National High-tech Research and Development Program of China (No. 2002BA711A12).

References

[1] H. Mihara, H. Sumi, T. Akazawa, et al., Fibrinolytic enzyme extracted from earthworm, Thromb. Haemostas. 50 (1983) 258–263.

[2] H. Mihara, H. Sumi, T. Yoneta, H. Mizumoto, R. Ikeda, M. Seiki, M. Maruyama, A novel fibrinolytic enzyme extracted from the earthworm, *Lumbricus rubellus*, Jpn. J. Physiol. 41 (1991) 461–472.

[3] N. Nakajima, H. Mihara, H. Sumi, Characterization of potent fibrinolytic enzymes in earthworm, *Lumbricus rubellus*, Biosci. Biotechnol. Biochem. 57 (1993) 1726–1730.

[4] N. Nakajima, M. Sugimoto, K. Ishihara, K. Nakamura, H. Hamada, Further characterization of earthworm serine proteases: cleavage specificity against peptide substrates and on autolysis, Biosci. Biotechnol. Biochem. 63 (1999) 2031–2033.

[5] F. Wang, C. Wang, M. Li, L. Gui, J. Zhang, W. Chang, Purification, characterization and crystallization of a group of earthworm fibrinolytic enzymes from *Eisenia fetida*, Biotechnol. Lett. 25 (2003) 1105–1109.

[6] Y. Tang, D. Liang, T. Jiang, J. Zhang, L. Gui, W. Chang, Crystal structure of earthworm fibrinolytic enzyme component a: revealing the structural determinants of its dual fibrinolytic activity, J. Mol. Biol. 321 (2002) 57–68.

[7] Y. Tang, J. Zhang, L. Gui, C. Wu, R. Fan, W. Chang, D. Liang, Crystallization and preliminary X-ray analysis of earthworm fibrinolytic enzyme component A from *Eisenia fetida*, Acta Crystallogr. D Biol. Crystallogr. 56 (Part 12) (2000) 1659–1661.

[8] T.A. Jones, J.Y. Zou, S.W. Cowan, Kjeldgaard, Improved methods for building protein models in electron density maps and the location of errors in these models, Acta Crystallogr. A 47 (Pt. 2) (1991) 110–119.

[9] A. Vagin, A. Teplyakov, MOLREP: an automated program for molecular replacement, J. Appl. Cryst. 30 (1997) 1022–1025.

[10] R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, PROCHECK: a program to check the stereochemical quality of protein structures, J. Appl. Crystallog. 26 (1993) 283–291.

[11] R.M. Stroud, A family of protein-cutting proteins, Sci. Am. 231 (1974) 74–88.

[12] J. Zhao, L. Li, C. Wu, R.Q. He, Hydrolysis of fibrinogen and plasminogen by immobilized earthworm fibrinolytic enzyme II from *Eisenia fetida*, Int. J. Biol. Macromol. 32 (2003) 165–171.

[13] L. Hedstrom, L. Szilagyi, W.J. Rutter, Converting trypsin to chymotrypsin: the role of surface loops, Science 255 (1992) 1249–1253.

[14] J.J. Perona, C.S. Craik, Structural basis of substrate specificity in the serine proteases, Protein Sci. 4 (1995) 337–360.