Multi-isomorphous replacement phasing of the earthworm fibrinolytic enzyme component A from *Eisenia fetida*

TANG Yong (汤 涌)[1], JIANG Tao (江 涛)[1], ZHANG Jiping (张季平)[1], FAN Rong (樊 蓉)[2], WU Cheng (吴 骋)[2], LIANG Dongcai (梁栋材)[1] & CHANG Wenrui (常文瑞)[1]

1. National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China;
2. Earthworm Fibrinolytic Enzyme Research Group, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China

Correspondence should be addressed to Chang Wenrui (email:wrchang@sun5.ibp.ac.cn)

Received September 17, 2002

Abstract Earthworm fibrinolytic enzyme component A (EFEa) from *Eisenia fetida*, a protein functioning not only as a direct fibrinolytic enzyme, but also as a plasminogen activator, has been crystallized in $P2_12_12_1$ space group with 3 protein molecules per asymmetric unit. Four heavy atom derivatives were prepared using a mother liquor containing 1.4 mol · L^{-1} Li_2SO_4 and 0.1 mol · L^{-1} MOPS buffer (pH7.2) and used to solve the protein's diffraction phase. The heavy atom binding sites in the derivative crystals were determined using difference Patterson and difference Fourier methods and were refined in combination to yield the initial protein's structure phase at 0.25 nm resolution. The non-crystallographic symmetry relationship of the three independent protein molecules in the asymmetric unit was determined using the correlative heavy atom sites and used for the averaging of the initial electron density. As a result, the electron density was significantly improved, providing a solid foundation for subsequent structure determination.

Keywords: *Eisenia fetida*, earthworm fibrinolytic enzyme, plasminogen, activator, multi-isomorphous replacement, phasing.

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, and eventually lead to thrombolysis. 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 streptokinase, have been extensively studied and used as thrombolytic agents in clinic. However, these agents are expensive and suffer from a number of significant limitations, such as fast clearance, reocclusion and bleeding complications[1]. Due to these limitations of the existing thrombolytic agents, novel thrombolytic agents with better fibrinolytic efficiency and less adverse side effects are eagerly sought.

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*[2] and *Eisenia fetida*[3] were characterized and became commercially available in Korea and

China as novel oral-administered fibrinolytic agents for prevention and treatment of cardiac and cerebrovascular diseases[4]1). 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 *Lumbricus rubellus* was composed of six serine protease components, of which two (F-III-1 and F-III-2) were trypsin-like, three (F-I-0, F-I-1, and F-I-2) were chymotrypsin-like, and the other component (F-II) was still unclear[5,6]. Among these components, F-II is very attractive because of its extraordinary substrate specificity and inhibition characteristics. F-II showed higher fibrinolytic and caseinolytic activities than other fibrinolytic components[5,6]. Moreover, F-II exhibited very strong proteolytic activity towards oxidized insulin B chains and β-amyloid 1-40, cleaving several peptide bonds behind basic, aromatic or hydrophobic residues[7]. However, by striking contrast, F-II was almost inactive with synthetic chromogenic substrates specific for trypsin, chymotrypsin or elastase[5,6]. F-II was significantly inhibited by soybean trypsin inhibitor (SBTI) and aprotinin, while its inhibition by tosyl-N-lysine chloromethyl ketone (TLCK), ε-aminocaproic acid (ε-ACA), or elastatinal (a strong elastase inhibitor) was very limited[5,6]. The reasons for all these observations have still not been identified.

Recently, we purified and crystallized a novel earthworm fibrinolytic enzyme from *Eisenia fetida* which we named earthworm fibrinolytic enzyme component A[8]. 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. EFEa was purified by a method similar to that for F-II[5]. Although the complete sequences of EFEa and F-II are still not available, the first 24 N-terminal amino acids of EFEa are identical to F-II[6]. The molecular weight of EFEa is essentially the same as F-II (24667 versus 24664, mass spectroscopy results)[6,8]. Furthermore, the unique incomplete fibrinolysis of EFEa as observed in the fibrin plate assay also resembles that of F-II[7]. These characteristics, together with the close evolutionary relationship2) of the two earthworm species *Eisenia fetida* and *Lumbricus rubellus*, enable us to infer that EFEa is extremely homologous, if not identical, to F-II.

In order to study the mechanism underlying EFEa's dual fibrinolytic activities and to locate the structural determinants of its extraordinary substrate specificity, a multi-isomorphous replacement (MIR) phasing was applied to solve the EFEa crystal structure. To this end, four heavy atom derivatives suitable for phasing were obtained after extensive screening. The heavy atom positions in the derivatives were determined and refined in combination to yield an initial structure phase at

1) Daedo Pamphlet, YONGSHIM capsule, Daedo Pharmaceutical Co., Ltd., 1990, 1—2.

2) see http://www.ncbi.nlm.nih.gov for the taxonomy.

0.25 nm resolution. The non-crystallographic symmetry relationship of the three independent protein molecules in the asymmetric unit was determined using the correlative heavy atom sites and used for the averaging of the initial electron density. As a result, the electron density was significantly improved, providing a solid foundation for subsequent structure determination.

1 Preparation of heavy atom derivatives and refinement of heavy atom parameters

The purification and crystallization of EFEa was carried out as described earlier[8]. 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 ($\alpha = \beta = \gamma = 90°$) (1 Å = 0.1 nm). There are three molecules per asymmetric unit, with a v/v solvent content of 45.0%[8].

When preparing heavy atom derivatives, a mother liquor was first found in which crystals could be stably preserved. Then, in principle, mercury compounds were dissolved at 2.0 mmol·L^{-1}, while other heavy atom reagents, such as platinum and gold compounds, were dissolved at 10.0 mmol·L^{-1} in the mother liquor. Crystals were transferred from the hanging drops to the mother liquor and soaked for 3 h before they were transferred to specific heavy atom solutions. The heavy atom concentrations were adjusted according to the changes of the soaked crystals. For the crystals that were well preserved in the heavy atom solutions, partial diffraction data were collected to evaluate the intensity changes on day 1, day 3 and day 6, when necessary. During the data collection of the heavy atom derivatives, partial diffraction data covering an oscillation angle of about 10 degrees were collected and compared with the native data set mentioned above. When the D_{iso} value (table 1) was higher than about 12% and the changes in the cell dimensions were less than 1%, data collection was continued to obtain a complete data set of the derivative crystal.

Table 1 The statistics of the native crystal and heavy atom derivatives diffraction data

Crystals	Native	C_2H_5HgCl	C_3H_7HgCl	$KAuCl_4$	K_2PtCl_6
Resolutions/nm	0.23	0.24	0.30	0.27	0.26
Images collected (× 1°)	270	202	140	180	240
No. of reflections	118978	87318	33286	51205	73808
Independent reflections	23546	20500	10630	16012	16930
Completeness[a] (overall/outmost shell) (%)	74/41	74/45	78/44	74/45	74/42
R_{sym}(on I)[b] (%)	9.5	10.3	10.9	12.0	10.2
D_{iso}(on F)[c] (%)	—	15.0	14.0	14.5	15.2

a) The completenesses were calculated with reflections with $I > 2.0 \times s_1$; b) $R_{sym} = \sum|I - \langle I \rangle| / \sum \langle I \rangle$; c) $D_{iso} = \sum(|Fph - Fp|) / \sum Fp$.

Since the reservoir solution used in crystallization, containing 2.0 mol·L^{-1} $(NH_4)_2SO_4$, 5.0% (v/v) PEG400, and 0.10 mol·L^{-1} MOPS buffer (pH 7.2), could well preserve crystals, it was first used as mother liquor to dissolve heavy atom reagents and screen for suitable heavy atom derivatives. To our disappointment, after screening over 50 heavy atom reagents, we only obtained two useful heavy atom derivatives with K_2PtCl_4 and $KAuCl_4$, respectively (data not shown). After

extensive screening of alternative mother liquor, a mother liquor containing 1.7 mol·L^{-1} Li_2SO_4 and 0.05 mol·L^{-1} MOPS (pH 7.2) was used as heavy atom soaking liquor to yield four useful heavy atom derivatives prepared with K_2PtCl_6, $KAuCl_4$, C_2H_5HgCl and C_3H_7HgCl, respectively. For the phasing of the heavy atoms, a native data set was also collected using crystal soaked in the above-mentioned mother liquor. X-ray diffraction data of the native crystal and all the heavy atom derivatives were collected at 18℃ using a MAR image plate system (MAR Research, Hamburg, Germany) at the State Key Lab of Biomacromolecules. The reflection intensities were integrated and scaled by DENZO and SCALEPACK[9]. The preparation conditions of the native crystal and its heavy atom derivatives are listed in table 2, while the statistics of their diffraction data is summarized in table 1.

Table 2 Heavy atom types and preparation conditions of the heavy-atom derivatives

Derivatives	Types and concentrations of the heavy atom reagents /mmol · L^{-1}		Crystal soaking time/h
Native[a)]	—	—	3
C2HG	C_2H_5HgCl	2.00	72
C3HG	C_3H_7HgCl[b)]	1.81	72
AUC4	$KAuCl_4$	1.30	72
PTC6	K_2PtCl_6	4.12	72

a) To ensure the reliability of the isomorphous differences, the native crystal used in the phase determination was also soaked in the above-mentioned mother liquor for 3 h before data collection; b) isopropyl mercury chloride.

The difference Patterson maps of the heavy atom derivatives were calculated using 0.90—0.30 nm data with the CCP4 program[1)]. The major heavy atom binding sites of the four heavy atom derivatives were determined using the Harker peaks on the Harker sections. The determination of the major heavy atom binding sites in the derivatives PTC6, AUC4 and C3HG using the difference Patterson method is illustrated in fig. 1(a), 1(b) and 1(c) (derivative C2HG shared the same major heavy atom binding sites with C3HG and was therefore not illustrated). Difference Fourier maps were calculated using the phases derived from the three major heavy atom binding sites of PTC6 as illustrated in fig. 1(a) and used to determined the minor heavy atom binding sites in derivative PTC6 and the major and minor heavy atom binding sites in all the other heavy atom derivatives. The origins of all the heavy atom coordinates were unified and the results from both the difference Patterson and difference Fourier maps were cross-validated. The heavy atom parameters of every heavy atom derivative were independently refined using MLPHARE[10] program. The refined parameters of all the heavy atom positions are listed in table 3.

1) Collaborative Computational Project Number 4 (CCP4), The CCP4 suite: Programs for protein crystallography, Acta Cryst., 1994, D50: 760—763.

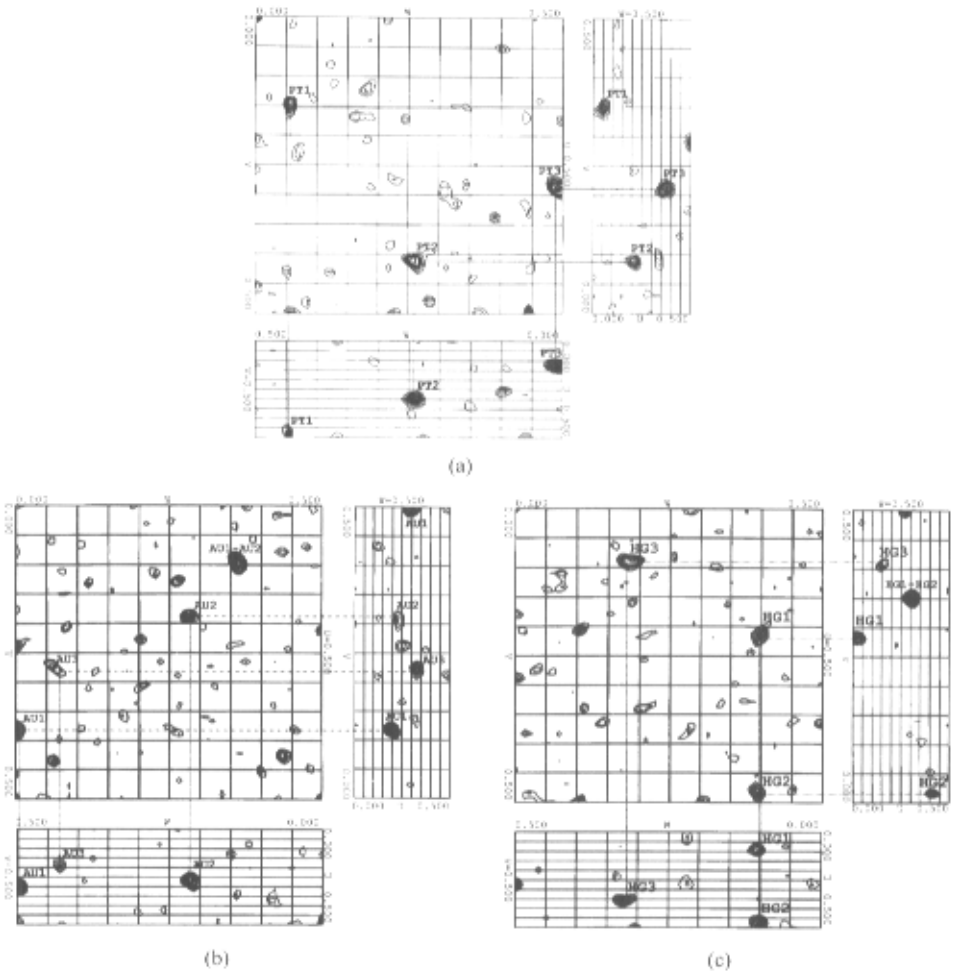

Fig. 1. The determination of the major heavy atom binding sites in the derivatives with difference Patterson method. The alignments of the Harker peaks of the three major heavy atom binding sites of derivatives PTC6, AUC4 and C3HG on the three Harker sections are shown in (a)—(c), respectively. Note that in (b) and (c), because of the special location of the heavy atom positions 1 and 2, their cross peaks also fell on the Harker sections. The labeling of the heavy atoms is the same as those in table 3.

2 Phase determination

The SIR phases obtained above were refined in combination to yield combined MIR phases. Although all the heavy atom derivatives diffract higher than 0.30 nm, for better reliability of the combined phases, the resolutions of the $KAuCl_4$ and K_2PtCl_6 derivatives data sets were truncated in the combined refinement according to the overall figure of Merit (FOM) of the reflections in different resolution ranges, as shown in table 4. The initial protein phases derived from the combined refinement were improved by solvent flattening and histogram matching with the DM[1)] program. The FOMs of the combined phases at 0.35 nm and 0.25 nm resolution were improved

1) See the footnote on page 266.

Table 3 The refined parameters of the heavy atom positions in the heavy-atom derivatives

Heavy atom types	Heavy atoms	Fractional coordinates			Occupancies[a)]	B factors (Å^2)
C_2H_5HgCl	HG1 [b)]	0.543	0.111	0.036	0.500	21.0
	HG2 [b)]	0.192	0.141	0.146	0.460	25.1
	HG3 [b)]	0.545	0.098	0.462	0.448	21.8
	HG4 [b)]	0.928	0.091	0.472	0.288	20.9
	HG5 [b)]	0.587	0.463	0.378	0.209	21.7
	HG6	0.654	0.298	0.423	0.139	33.2
	HG7	0.166	0.208	0.158	0.103	16.9
C_3H_7HgCl	HG1	0.543	0.111	0.036	0.393	18.4
	HG2	0.192	0.141	0.146	0.360	24.0
	HG3	0.545	0.098	0.462	0.373	22.6
	HG4	0.928	0.091	0.472	0.054	17.2
$KAuCl_4$	AU1 [b)]	0.033	0.340	0.306	0.372	26.1
	AU2 [b)]	0.333	0.437	0.376	0.361	18.5
	AU3 [b)]	0.680	0.236	0.350	0.245	10.9
	AU4	0.236	0.382	0.148	0.085	22.0
K_2PtCl_6	PT1 [b)]	0.319	0.212	0.033	0.565	29.1
	PT2 [b)]	0.367	0.388	0.437	0.536	28.0
	PT3 [b)]	0.147	0.150	0.077	0.556	31.6
	PT4	0.290	0.041	0.243	0.082	26.6

a) The occupancies within the resolution limits used in the refinement (table 4); b) the heavy atom positions used in the determination of the NCS relationship of the three independent protein molecules in the asymmetric unit.

from 0.618 and 0.446 (table 4) to 0.871 and 0.811, respectively by the DM density modification, indicating the reliability of model building at these two resolution levels. DM-improved electron density maps were calculated at 0.35 nm, 0.30 nm and 0.25 nm resolutions for subsequent use in structure determination.

Table 4 Refinement of heavy atom positions and calculation of protein phase

Crystals		Native	C_2H_5HgCl	C_3H_7HgCl	$KAuCl_4$	K_2PtCl_6
Heavy atom binding sites		—	7	4	4	4
Resolution used in refinement /nm		0.23	0.25	0.30	0.41	0.35
Acentric reflections	Phasing power [a)]	—	1.72	1.47	1.28	1.34
	R_{cullis} [b)]	—	0.73	0.78	0.82	0.82
	FOM (SIR)	—	0.311	0.287	0.265	0.274
Centric reflections	Phasing power [a)]	—	1.39	1.18	1.05	1.11
	R_{cullis} [b)]	—	0.59	0.64	0.70	0.69
	FOM (SIR)	—	0.564	0.515	0.501	0.519
FOM (all reflections)		—	0.337	0.314	0.302	0.308
FOM (combined, 0.35 nm)			$FOM_{accentric}$= 0.585, $FOM_{centric}$= 0.811, FOM_{all}= 0.618			
FOM (combined, 0.25 nm)			$FOM_{accentric}$= 0.415, $FOM_{centric}$= 0.700, FOM_{all}= 0.446			

a) Phasing power= $[\sum|Fh|^2/\sum(|Fph(obs)| - |Fph(calc)|)^2]^{1/2}$; b) $R_{cullis}= \sum||Fph\pm Fp| - Fph(calc)|/\sum|Fph\pm Fp|$.

3 Determination of the non-crystallographic symmetry and improvement of the electron density maps

The non-crystallographic symmetry (NCS) relationship of the three independent molecules in the asymmetric unit was determined using the correlative heavy atom positions with the Site2RT program in RAVE[10] and used to generate a molecular envelope around molecule A using MAMA[10]. The matrices were further improved by IMP[10], with which a more precise molecule envelope was calculated and used in the NCS averaging of the electron density maps mentioned above with SPANCSI program[10].

The heavy atom positions being used to determine the NCS relationship were marked in table 3 and illustrated in fig. 2. The location and distances of the heavy atoms indicated that they have bound to the same positions of the three independent protein molecules (referred to as A, B and C), respectively. Thus the NCS relationship between molecule A and B and A and C could be calculated with 3 and 4 pairs of heavy atom coordinates, respectively. As a result, the three independent molecules in the asymmetric unit were found to be related by two non-crystallographic two-fold axes. These two axes are substantially parallel to bc plane (result not shown) and form an angle of about 30° with each other as viewed along the crystallographic axis a. The axis relating mole-

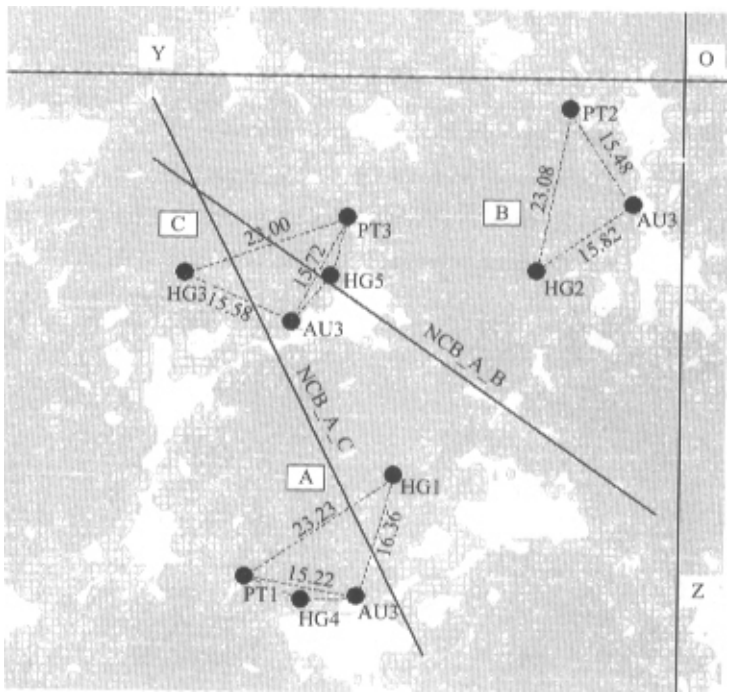

Fig. 2. The application of the correlative heavy atom positions in the determination of the NCS relationship of the three independent molecules. The location and distances of the heavy atoms indicated that they have bound to the same positions of the three independent protein molecules (referred to as A, B and C), respectively. This made the NCS relationship between molecule A and B and A and C could be calculated with 3 and 4 pairs of heavy atom coordinates, respectively. The non-crystallographic two-fold axes relating molecules A and B and A and C were labeled as NCS_A_B and NCS_A_C, respectively. The NCS-averaged electron density at 0.25 nm resolution is shown in the cell (in gray) to illustrate the clear protein-solvent boundary and the NCS relationship of the electron density of the protein molecules.

cules A and B is substantially proper[11], while the one relating molecules A and C is improper[11]. The correlation coefficients of the electron density of molecules A and B and molecules A and C related by the improved NCS matrices were 0.75 and 0.74, respectively, indicating the good NCS relationship (fig. 3). The existence and orientations of the two non-crystallographic two-fold axes in the asymmetric unit were in consistence with the results of the self-rotation function calculated with the native data set.

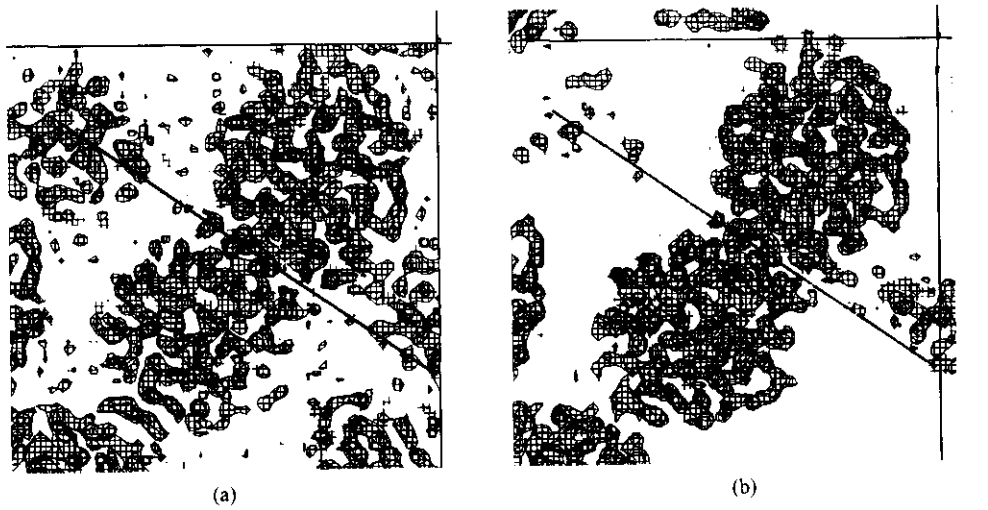

Fig. 3. The non-crystallographic two-fold axis symmetry of molecule A and B and the improvement of their local electron density by NCS averaging. (a) The DM modified electron density map at 0.35 nm resolution; (b) the DM-modified and NCS-averaged electron density map at 0.25 nm resolution.

The NCS averaging significantly improved the overall quality of the electron density, making the protein-solvent boundary clearly identifiable (figs. 2 and 3) and the trace of the peptide chain unambiguous (fig. 4). The good-quality electron density has enabled us to trace the peptide chain

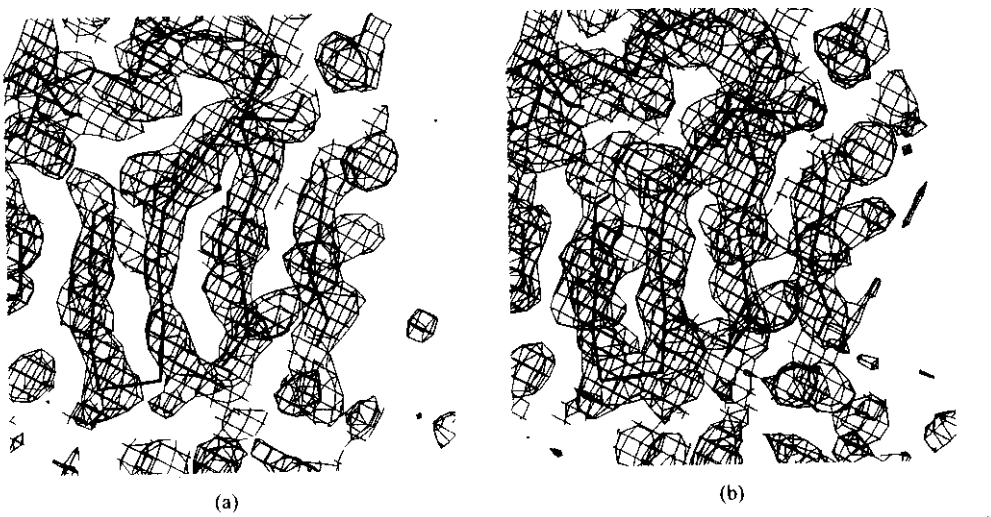

Fig. 4. The unambiguous trace of the EFEa molecule's peptide chain (partial) in (a) the DM-modified electron density map at 0.35 nm resolution, and (b) the DM-modified and NCS-averaged electron density at 0.25 nm resolution. In the figures, the traces of the peptide chains were illustrated by the lines connecting the initially determined Cα positions.

and fit the side chains without much difficulty. EFEa molecule was initially identified as a globular protein with a diameter of about 2.5 nm and is composed of many β sheets and few α helixes.

4 Discussion

EFEa was crystallized with a highly concentrated ammonium sulfate solution. Since crystals kept stable in the reservoir solution used for crystallization, the reservoir solution was used as the heavy atom soaking liquid in the early stage of heavy atom derivative preparations. Although these two heavy atom derivatives had provided us with initial structure phases, more heavy-atom derivatives had to be prepared for the improvement of the structure phases reliability and accuracy. The high salt concentration and the ammonia dissociated from the highly concentrated ammonium sulfate were assumed to be responsible for the unsuccessful binding of the heavy atom reagents with the EFEa enzyme. Therefore, a suitable alternative mother liquor must be found to prepare more derivatives. After extensive screening, the above-mentioned lithium sulfate mother liquor was found suitable and then used to prepare the four useful heavy atom derivatives as mentioned above in quite a short time.

The NCS relationship of the three independent molecules in the asymmetric unit was successfully determined using the heavy atom positions and was used in the averaging of the electron density, which significantly improved the electron density quality. During the NCS averaging of the electron density, a molecular envelope larger than the protein-solvent boundary of molecule A was made so as to include the electron density of molecule A as much as possible. Although the NCS averaging improved the local electron density for the main part of the molecules, it deteriorated the electron density in some flexible regions, such as the solvent-exposed regions. Therefore, the DM modification and NCS averaging of the electron density were performed separately. Thus, in model building, the peptide chain tracing and the residue fitting, especially the fitting of the side chains, could be undertaken advantageously with reference to two different kinds of electron density maps and on the basis of rational analysis. The advantage of calculating and using electron density maps at different resolution levels is that these maps might be complementary in some uncertain regions. In addition to the successful application in the electron density averaging, the determination of the NCS relationship would also be expected to facilitate the subsequent model building and structural refinement. Once the EFEa structure is solved, it will be the first structure of an earthworm fibrinolytic enzyme component and of a serine protease originating from the annelid worms. We expect that the solution of the EFEa crystal structure will provide us with important structural information for the understanding of the molecular mechanism underlying the enzyme's biological function, especially the dual fibrinolytic activity and for further pharmaceutical application and rational drug design of the enzyme.

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

References

1. Lijnen, H. R., Collen, D., Fibrinolytic agents: Mechanisms of activity and pharmacology, Thromb. Haemostas., 1995, 74: 387–390.
2. Mihara, H., Sumi, H., Akazawa, K. et al., Fibrinolytic enzyme extracted from the earthworm, Thromb. Haemostat., 1983, 50: 258.
3. Zhou, Y. C., Zhou, H., Chen, Y. C., Purification and biochemical characterization of the fibrinolytic enzymes from the earthworm *Eisenia fetida*, Acta Biochem. Biophys. Sinica (in Chinese), 1988, 20: 35—42.
4. Hou, Q. M., Mass production of earthworm fibrinolytic enzyme under GMP standard, Prog. Biochem. Biophys. (in Chinese), 1995, 22(6): 567.
5. Mihara, H., Sumi, H., Yoneta, T. et al., A novel fibrinolytic enzyme extracted from the earthworm *Lumbricus rubellus*, Japa. J. Physiol., 1991, 41: 461—472.
6. Nakajima, N., Mihara, H., Sumi, H., Characterization of potent fibrinolytic enzymes in earthworm *Lumbricus rubellus* Biosci. Biotech. Biochem., 1993, 57: 1726—1730.
7. Nakajima, N., Sugimoto, M., Ishihara, K. et al., Further characterization of earthworm serine proteases: Cleavage specificity against peptide substrates and on autolysis, Biosci. Biotechnol. Biochem., 1999, 63(11): 2031—2033.
8. Tang, Y., Zhang, J., Gui, L. et al., Crystallization and preliminary X-ray analysis of earthworm fibrinolytic enzyme component A from *Eisenia fetida*, Acta Cryst., 2000, D56: 1659—1661.
9. Otwinowski, Z., Minor, W., Processing of X-ray diffraction data collected in oscillation mode, Methods Enzymol., 1997, 276: 307—326.
10. Kleywegt, G. J., Jones, T. A., Software for handling macromolecular envelopes, Acta Cryst., 1999, D55: 941—944.
11. Vellieux, F. M. D., Read, R. J., Noncrystallographic symmetry averaging in phase refinement and extension, Method in Enzymol., 1997, 277: 18—53.