Proc. Natl. Acad. Sci. USA
Vol. 74, No. 2, pp. 556–559, February 1977
Biochemistry

Homology among acid proteases: Comparison of crystal structures at 3 Å resolution of acid proteases from *Rhizopus chinensis* and *Endothia parasitica*

(x-ray diffraction/inhibitor binding/enzyme cleft)

E. SUBRAMANIAN*, I. D. A. SWAN*†, MAMIE LIU*, D. R. DAVIES*, J. A. JENKINS‡§, I. J. TICKLE‡§, AND T. L. BLUNDELL‡§

* Laboratory of Molecular Biology, National Institute of Arthritis, Metabolism, and Digestive Diseases, Bethesda, Maryland 20014; and ‡ Biochemistry Laboratory, School of Biological Sciences, Sussex University, Falmer, Brighton BN19QG, England

Communicated by Gary Felsenfeld, December 9, 1976

ABSTRACT **The molecular structures of two fungal acid proteases at 3 Å resolution have been compared, and found to have similar secondary and tertiary folding. These enzymes are bilobal and have a pronounced cleft between the two lobes. This cleft has been identified as the active site region from inhibitor binding studies. The results of the comparison are discussed in terms of homology among the acid proteases in general.**

The acid proteases form an interesting and well-defined class of enzymes. They include the mammalian enzymes pepsin and chymosin (rennin), and a number of fungal enzymes such as those from *Rhizopus chinensis*, *Endothia parasitica*, and *Penicillium janthinellum* (see refs. 1 and 2). These enzymes are all characterized by molecular weights in the region of 35,000, and pH optima for catalytic action in the range of 1.5–5.0. They share the property of being inhibited by epoxy compounds and by diazonium compounds in the presence of Cu^{2+} ions (3, 4). Their catalytic activities have been studied extensively and shown to involve two aspartic acid residues [Asp 32 and Asp 215 in the pepsin sequence (5)]. They are all inhibited by a microbial oligopeptide, pepstatin, an acid-protease-specific inhibitor obtained from culture filtrates of *Streptomyces* (6–9). They all possess essentially similar substrate specificity, preferring hydrophobic residues on both sides of the scissile bond. In addition, their catalytic action appears to involve extended subsite interactions with several amino acid side chains of the substrate (1, 2).

The complete amino acid sequence has been determined for porcine pepsin (5), and limited sequence information is available for several other acid proteases (10–12). These data reveal similarity in sequence and suggest there is homology among the various acid proteases. In particular, the residues immediately adjacent to the two catalytically important aspartic acid residues appear to be identical in all the cases that have been examined (see ref. 1).

X-ray diffraction analyses are in progress for several of the acid-proteases and preliminary low-resolution maps have been presented for pepsin at 5 Å (13), the *Endothia parasitica* acid protease at 5.2 Å (14), the *Rhizopus chinensis* enzyme at 5.5 Å (15), and the *Penicillium janthinellum* enzyme at 6 Å (16). Here we present the preliminary results of two independent crystal structure analyses at 3.0 Å resolution of the enzymes from *Rhizopus chinensis* and *Endothia parasitica*, together with a comparison of their structures.

X-ray structure determination

The crystal structures of the acid proteases from *Rhizopus chinensis* and *Endothia parasitica* have been determined. MIR (multiple isomorphous replacement) methods were used for the *Rhizopus* enzyme, and MIR plus anomalous scattering was used for that from *Endothia parasitica*. From electron-density maps, practically the entire peptide backbone of the single polypeptide chain has been traced for both enzymes. A brief description of the structure analysis for each enzyme follows.

The Crystal Structure of the Acid Protease from *Rhizopus chinensis*. The enzyme was isolated and crystallized ten years ago by Fukumoto *et al.* (17). The enzyme used here was obtained commercially from Miles Laboratories. Conditions for obtaining crystals suitable for x-ray analysis have been described (18). The crystals are orthorhombic, spacegroup $P2_12_12_1$, with unit cell dimensions of a = 60.3, b = 60.7, and c = 107.0 Å. There is one molecule per asymmetric unit and the solvent content is about 56%.

The low-resolution x-ray analysis (15) showed clearly that the molecule was bilobal and that there was a large cleft between the lobes. A 3 Å electron-density map (figure-of-merit = 0.78) was calculated for the enzyme using the heavy-atom derivatives shown in Table 1. Because the amino acid sequence is not available, we were dependent on the continuity of the electron density to trace the polypeptide backbone, and most of the backbone can be followed in this way. A large number of the amino acid side chains are clearly visible. In a few places, however, particularly at bends in the backbone, there are breaks in the density that result in some uncertainties in the chain connectivity. The chain tracing is also somewhat equivocal near the base of the cleft, where the two lobes come in close contact and there are several bridges in the density that presumably result from hydrogen bonding. Nevertheless, the continuity is generally quite reasonable and provides a good overall picture of the molecular configuration. Three-hundred twenty-four α-carbon positions could be assigned; the amino acid composition also suggests 324 residues. The known sequence for the NH_2-terminal 39 residues (11, 12) could be fitted to an end of the polypeptide chain that was therefore assigned to be the NH_2-terminus (Fig. 1a). This assignment locates the catalytically active Asp 35 in a loop at the base of the cleft. Adjacent to this loop, but attached to the other lobe on the other side of the cleft, is a similar loop containing the residues around 215. It is interesting to note that Asp 215 is the other catalytically active residue in pepsin.

Fig. 1a is a stereo drawing of the α-carbon backbone of the *Rhizopus* enzyme. The molecule is mostly made up of antiparallel pleated sheet structure with little helix. The prominent

† Present address: Department of Chemistry, University of Glasgow, Glasgow, Scotland.
§ Present address: Laboratory of Molecular Biology, Department of Crystallography, Birkbeck College, London University, Malet Street, London WCIE 7HX, England.

Table 1. Heavy-atom derivatives used in calculating 3 A electron-density map for *Rhizopus chinensis* acid protease

Derivative	Number of sites	Resolution of data, A	RMS(E)/RMS(F_H)	R_c	R_k
Pb acetate (Pb)	3	3.0	0.494	0.505	0.080
Uranyl acetate	6	3.0	0.499	0.551	0.108
Uranyl fluoride (UF)	5	4.0	0.774	0.546	0.111
Pb + UF	3	3.0	0.556	0.555	0.081
Merbromin	3	4.0	0.920	0.629	0.080
Iodine	13	3.0	0.539	0.536	0.112
Pb + Iodine	18	3.0	0.517	0.575	0.154
Iridium	1	4.5	0.721	0.539	0.064
Biilmann's reagent*	4	5.5	0.506	0.527	0.081
Baker's reagent†	6	5.5	0.563	0.575	0.132

RMS is root mean square; E is the lack-of-closure error of the phase triangle. Figure of merit at 3.0 Å = 0.78.

$$R_c = \frac{\Sigma||F_{PH}| - |F_P| - |F_H||}{\Sigma||F_{PH}| - |F_P||} \qquad R_k = \frac{\Sigma||F_{PH}| - |F_P + F_H||}{\Sigma|F_{PH}|}$$

For R_c, the summation is over centrosymmetric reflections only. For R_k, the summation is over entire reflections. F_p is the structure amplitude for the parent enzyme, F_{PH} is for the derivative enzyme, and F_H is for the heavy-atom contribution alone.

* $C_6H_{10}O_2Hg_2SO_4 \cdot H_2O$.

† $[CH_3COOHgCH_2CH(OCH_3)]_2$.

cleft between the two lobes has been identified as the active site region from a study of the binding of several inhibitors to the enzyme in the crystal. The most interesting results were obtained with the microbial peptide pepstatin, whose structure (7) is shown in Fig. 2. It has an unusual amino acid, statine, is essentially a hexapeptide, and in a fully stretched conformation could have a length of about 20 Å. Crystals of the enzyme were soaked in a solution of saturated pepstatin for several weeks, at the end of which a complete set of 3.7 Å data was collected from one crystal. A difference electron density map, calculated using the final phases, showed clearly that the pepstatin molecule binds in the cleft region. It appears to be in an extended conformation, with a length of about 20 Å. These results are consistent with proposals from kinetic studies (2) of an extended binding site for acid proteases.

The Crystal Structure of the Acid Protease from *Endothia parasitica*. The protease was extracted from commercially available rennet "Sure Curd" (kindly donated by J. L. Sardinas of Pfizer Chemical Co.) by the method of Whitaker (19). Crystals of spacegroup $P2_1$, with cell dimensions of a = 53.6, b = 74.05, and c = 45.7 Å, $\beta = 110°$ were grown from 0.1 M sodium acetate adjusted to pH 6.3 and containing 2.7 M ammonium sulfate and 10% by weight of acetone (20). There is one molecule per asymmetric unit and the solvent content is about 49%. Details of a low-resolution (5.2 Å) electron density map have been published (14), and showed that the molecule is organized in two globular domains with a well-defined cleft.

A 3 Å electron density map was calculated for the enzyme using the six heavy atom derivatives shown in Table 2. The mean figure of merit is 0.70 for about 6600 reflections. The electron density allows a straightforward identification of the molecular boundary and there are large regions of solvent. Within the molecular boundary the polypeptide chain can be followed and many large and well-defined side chains are apparent. The chain can be followed for about 305 amino acid residues although there are two regions of ambiguity that in the absence of sequence data are difficult to resolve unequivocally. The first is a break in the continuity near one terminus and on the surface of the molecule. This may be due to errors in phasing, to disorder of the chain, or possibly to autolysis. The second relates to electron density near the center of the molecule at the base of the deep cleft. Four different chains converge, and there are clearly hydrogen bonds between the main chains; in this region the chain connectivity is ambiguous.

Fig. 1b shows the positions of the α-carbon atoms along with the most probable chain connectivity. There are two well-defined lobes of the molecule, with a deep cleft between them.

Table 2. Heavy-atom derivatives used in calculating the 3 A electron-density map for *Endothia parasitica* acid protease

Derivative	Number of sites	Range in $\sin^2\theta/\lambda^2$	$\bar{F}_H$	E	$\bar{F}''$	E''	$\|F_{PH} - F_P\|$
Uranyl acetate	5	0.0000–0.0096	110	48	20	32	86
		0.0096–0.0192	83	38	17	35	63
		0.0192–0.0288	64	46	13	69	71
K_2PtCl_4	1	0.0000–0.0096	77	47	—	—	72
		0.0096–0.0192	64	48	—	—	62
		0.0192–0.0288	—	—	—	—	—
$Pt(en^*)Cl_2$	2	0.0000–0.0096	65	47	—	—	67
		0.0096–0.0192	49	40	—	—	52
		0.0192–0.0288	36	58	—	—	59
K_2HgBr_4	2	0.0000–0.0096	80	56	9	28	76
		0.0096–0.0192	54	66	5	41	69
		0.0192–0.0288	37	57	4	44	64
$KAu(CN)_2$	3	0.0000–0.0096	93	72	—	—	97
		0.0096–0.0192	76	72	—	—	79
		0.0192–0.0288	—	—	—	—	—
KI_3	7	0.0000–0.0096	108	68	—	—	97
		0.0096–0.0192	94	85	—	—	115
		0.0192–0.0288	85	89	—	—	120

Range in $\sin^2\theta/\lambda^2$	$\bar{F}_P$	Number of reflections	Figure of merit
0.0000–0.0096	370	1339	0.89
0.0096–0.0192	407	2503	0.74
0.0192–0.0288	292	2758	0.57

The Fs have the same meaning as in Table 1. The double-prime refers to contribution from anomalous scattering.

* Ethylenediamine.

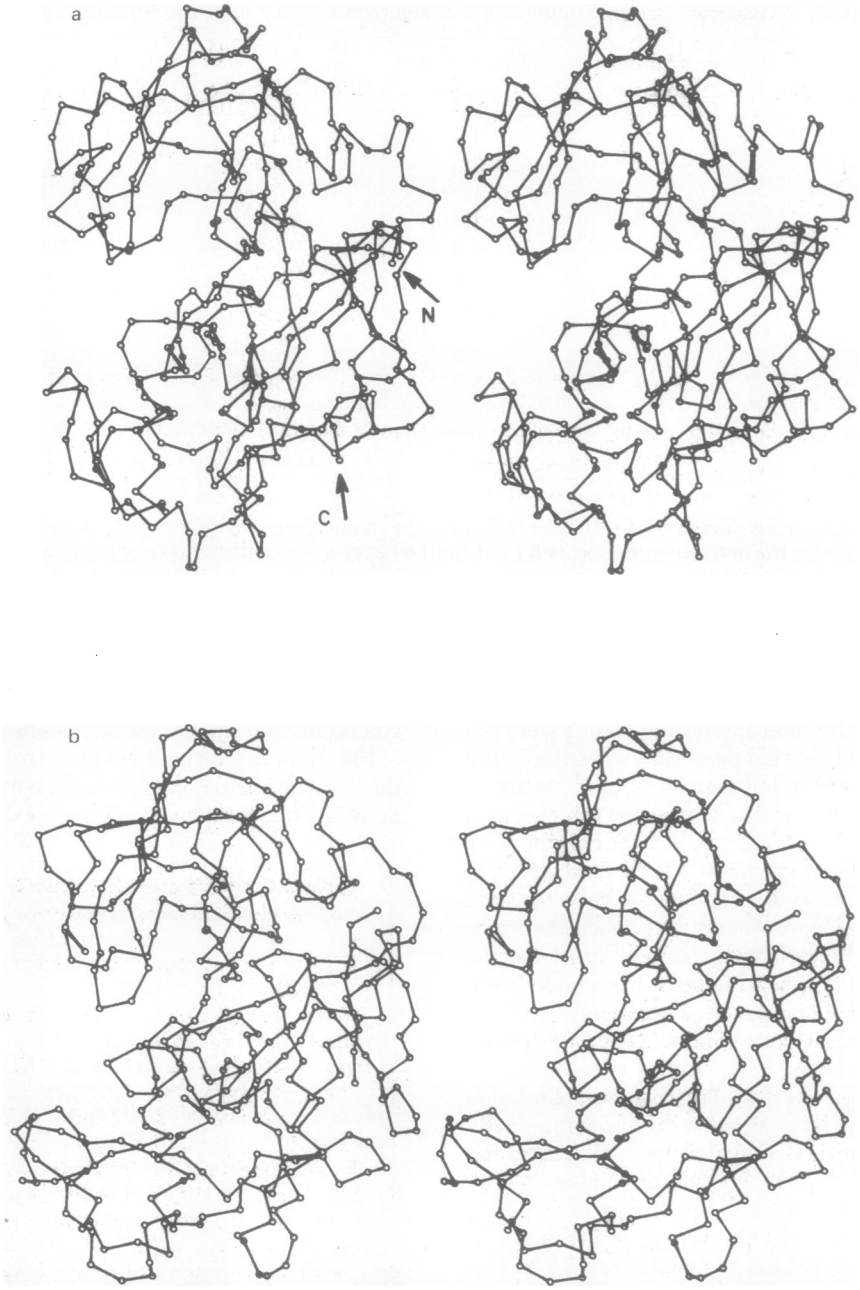

FIG. 1. Stereo drawings of the α-carbon chain tracing for the acid proteases from (a) *Rhizopus chinensis* and (b) *Endothia parasitica*. The molecules are bilobal with their active sites located in the cleft between the lobes. They are constructed almost entirely of extended segments of antiparallel β-structure. A frequent feature is the presence of loops connecting pairs of antiparallel strands. N and C indicate NH_2- and COOH-termini.

The chain forms many sheet-like structures with well-defined loops and very little helix.

Comparison of the two enzymes

From an examination of Fig. 1, it is apparent that the two enzymes closely resemble one another, not only in being bilobal with a pronounced cleft, but also in having almost identical secondary and tertiary structures. Both molecules have the approximate dimensions of 37 × 46 × 61 Å. The two lobes have contacts over a large area, giving rise to a broad waist rather similar to that found in papain (21), but in contrast to the narrow region between the two lobes in phosphoglycerate kinase (22, 23). The polypeptide chain first completes one lobe and then crosses over to complete the second, COOH-terminal lobe. The cleft between the lobes has a length of about 25 Å, and from the pepstatin binding studies on the *Rhizopus* enzyme, the cleft region can be identified as the site of catalytic action.

The lower COOH-terminal lobe of the molecules in Fig. 1 consists of about 11 extended strands arranged mostly antiparallel to one another, which together with one short helical region are wrapped around what appears to be a hydrophobic core.

The upper lobe is also constructed almost entirely of extended segments, mostly arranged in an anti-parallel fashion. It comprises three sheets stacked on top of each other, each containing three strands, with the direction of the strands rotated by about 90° in going from one sheet to another. This lobe includes a well-defined pocket, in the vicinity of Asp 35, possibly a specificity pocket, located at one side of the cleft. In both enzymes this pocket contains a site of iodination, and the presence of

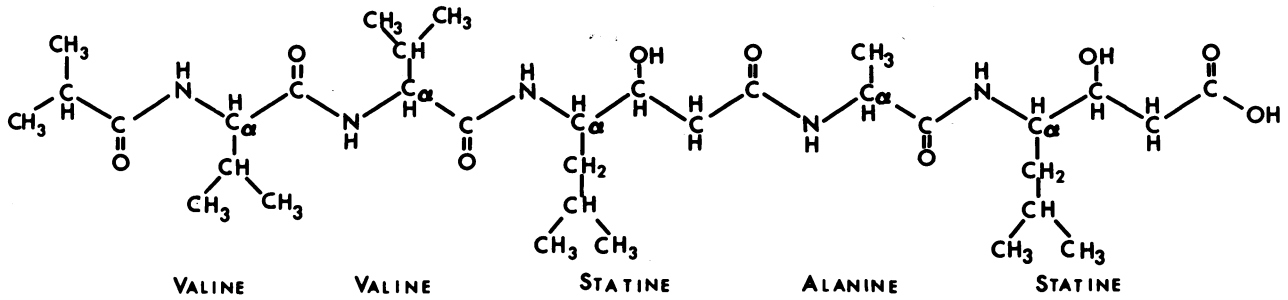

FIG. 2. The chemical structure of pepstatin (7).

what appear to be large aromatic side chains suggests a hydrophobic environment.

There are only four short helical regions. Three of these lie on the surface of the molecule and away from the cleft; the fourth is located about 20 residues away from the COOH-terminus. In both enzymes these helices contain from 1½ to 2½ turns. One of the helices, containing about 2 turns and observed in the low-resolution map for the *Rhizopus* enzyme (15), is noticeably shorter in the *Endothia* enzyme.

The fact that two independent investigations have produced the same connectivity and similar tertiary structures, even in the absence of sequence data, strengthens the conclusion that the two molecules are very closely related.

Where there are sequence data available for comparative purposes, as in pepsin (5), chymosin (10), penicillopepsin (12), and the *Rhizopus* enzyme (11, 12), it is apparent that homologies exist. For instance, there is as much as 50% identity between the first 38 residues of the *Rhizopus* enzyme and porcine pepsin (12). These data too strongly support the generalization that all these acid proteases are likely to be homologous, and will therefore have the same structure as that reported here.

Note Added in Proof. The structure of penicillopepsin has been determined at 2.8 Å resolution (24). A visual comparison indicates that this enzyme has a structure similar to that of the *Rhizopus* and *Endothia* enzymes.

We are grateful to Dr. G. H. Cohen and R. J. Feldmann for help with display programs on an Adage Graphic Terminal, which enabled us to visualize the homology between the two molecules. J.A.J., I.J.T., and T.L.B. would like to thank Dr. C. Bunn and Dr. P. Moews for making available data from their preliminary studies on the enzyme from *E. parasitica* and Dr. Luciano Ungaretti for helpful discussions; and the Science Research Council (United Kingdom) for financial support. We wish to thank Dr. T. Hofmann for some interesting discussions on the comparison between the *Rhizopus* and *Penicillum* enzymes.

1. Hofmann, T. (1974) "Structure, function and evolution of acid proteases," in *Advances in Chemistry Series, Number 136, Food Related Enzymes* (American Chemical Society, Washington, D.C.), pp. 146–185.
2. Fruton, J. S. (1976) *Adv. Enzymol.* **44**, 1–36.
3. Tang, J. (1971) *J. Biol. Chem.* **246**, 4510–4517.
4. Rajagopalan, T. G., Stein, W. H. & Moore, S. J. (1966) *J. Biol. Chem.* **241**, 4295–4297.
5. Tang, J., Sepulveda, J., Marciniszyn, K. C. S., Jr., Chen, W. Y., Huang, N. T., Liu, D. & Lanier, J. P. (1973) *Proc. Natl. Acad. Sci. USA* **70**, 3437–3439.
6. Aoyagi, T., Kunimoto, S., Morishima, H., Takeuchi, T. & Umezawa, H. (1971) *J. Antibiot.* **24**, 687–694.
7. Umezawa, H., Aoyagi, T., Morishima, H., Matsuzaki, M., Hamada, M. & Takeuchi, T. (1970) *J. Antibiot.* **23**, 259–262.
8. Aoyagi, T., Morishima, H., Nishizawa, R., Kunimoto, S., Takeuchi, T. & Umezawa, H. (1972) *J. Antibiot.* **25**, 689–694.
9. Takahashi, K. & Chang, W. J. (1973) *J. Biochem.* **73**, 675–677.
10. Pedersen, V. B. & Foltmann, B. (1975) *Eur. J. Biochem.* **55**, 95–103.
11. Sepulveda, P., Jackson, K. W. & Tang, J. (1975) *Biochem. Biophys. Res. Commun.* **63**, 1106–1112.
12. Gripon, J.-C., Rhee, S. H. & Hofmann, T. (1977) *Can. J. Biochem.*, in press.
13. Andreeva, N. S., Borisov, V. V., Melik-Adamyan, R. R., Raiz, V. Sh., Trofimova, L. N. & Shutskever, N. E. (1971) *Mol. Biol. USSR* **5**, 908–916.
14. Jenkins, J. A., Blundell, T. L., Tickle, I. J. & Ungaretti, L. (1975) *J. Mol. Biol.* **99**, 583–590.
15. Subramanian, E., Swan, I. D. A. & Davies, D. R. (1975) *Biochem. Biophys. Res. Commun.* **68**, 875–880.
16. Hsu, I.-N., Hofmann, T., Nyburg, S. C. & James, M. N. G. (1976) *Biochem. Biophys. Res. Commun.* **72**, 363–368.
17. Fukumoto, J., Tsuru, D. & Yamamoto, T. (1967) *Agr. Biol. Chem.* **31**, 710–717.
18. Swan, I. D. A. (1971) *J. Mol. Biol.* **60**, 405–407.
19. Whitaker, J. R. (1970) in *Methods in Enzymology*, eds. Perlmann, G. E. & Lorand, L. (Academic Press, New York), Vol. 19, pp. 436–445.
20. Moews, P. & Bunn, C. W. (1970) *J. Mol. Biol.* **54**, 395–397.
21. Drenth, J., Jansonius, J. N., Koekoek, R., Swen, H. M. & Wolthers, B. G. (1968) *Nature* **218**, 929–932.
22. Bryant, T. N., Watson, H. C. & Wendell, P. L. (1974) *Nature* **247**, 14–17.
23. Blake, C. C. F. & Evans, P. R. (1974) *J. Mol. Biol.* **84**, 585–601.
24. Hsu, I., Delbaere, L. T. J., James, M. N. G. & Hofmann, T. (1977) *Nature*, in press.