Vol. 80, No. 4, 1978 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
February 28, 1978 Pages 891-896

# THE STRUCTURE OF ACID PROTEASE FROM ENDOTHIA PARASITICA IN CROSS-LINKED FORM AT 3.5 A RESOLUTION

C. Wong, T.J. Lee, T.Y. Lee, T.H. Lu & C.S. Hung

National Tsing Hua University, Hsinchu, Taiwan, China

Received October 4, 1977

Summary: Electron density map at 3.5 A resolution has been prepared for the strongly cross-linked acid protease from Endothia parasitica with phase angles determined by three uranyl and one platinite derivatives. The mean figure of merit is 0.85. The map fits the known acid protease foldings, and accounts for 326 residues.

The importance and historical background for x-ray structural determination of acid proteases have recently been presented elaborately (1,2). We shall not repeat here.

In the course of our investigation of the structure of acid protease from Endothia parasitica, we are convinced, after careful and thorough scrutiny, that the strongly cross-linked protease crystal (3) is suitable for high resolution x-ray structural study. The following is the description of our investigation.

Experimental: The crude protease, a gift from Dr. Sardinas of Pfitzer Chem. Co., was purified by a modified Whitaker method (4,5). Crystals were grown as described by Moews and Bunn (6), except the pH was adjusted to 6.8 and without seeding. Largest crystals obtained had the size of 1x3x0.5mm, and the average was 0.5x1x0.25mm. The crystals were cross-linked with glutaraldehyde (3).

The heavy atom derivatives of three uranyls and one $K_2PtCl_4$ (UA, UB, UC & Pt) were prepared as described in our previous report (3). Other heavy atoms that gave intensity changes were $ThCl_4$($10^{-3}$M, 7 days) and $K_2HgBr_4$ ($10^{-2}$M, pH=5, 24 hours). However, they were not included in the cal-

0006-291X/78/0804-0891$01.00/0

culation because the strength of these two derivatives extended only to 5A.

X-ray data for the cross-linked (CL) and untreated (NT) natives and all the derivative crystals were collected on a Syntex $P\bar{1}$ autodiffractometer, using Cu $K_\alpha$ radiation. $\omega$-scan was employed with the background measured before and after each peak counts. Three reflections were chosen to monitor the decay of the protein crystal. The crystals were discarded when the checking reflections showed a loss of 10% of their original counts. The data were corrected for Lp and absorption (7) factors. Friedel pairs were collected for the natives (CL & NT) and all the derivatives mentioned above. Friedel pair differences $(R(+,-)=2\sum||F(+)|-|F(-)||/\sum||F(+)|+|F(-)||)$ for the native (CL & NT) are both 0.037 (3.5A), the uranyls (UA, UB, UC) 0.055-0.060 (3.5A) and Pt, 0.035 (5.0A). The Friedel pairs were kept in separation for the uranyl derivatives. The merging R factors for the F's were between 0.035-0.045.

UA difference Patterson was first solved with Harker section and an $ok\ell$ projection (2 major sites). The protein phase determined by UA isomorphous replacement and anomalous scattering data were used to find UB, UC and Pt heavy atom positions and their relative origins with respect to UA as well as the correct enantiomorph of the protein. The coordinates, occupancy and temperature factors were first refined by least square on each uranyl derivative individually with empirical form factor. The coefficient $F_{HLE}$(8,9,10) was used. After reaching convergence, difference maps (major sites substracted) were calculated for the uranyls to locate the minor sites. Several cycles of phase refinement (11,12) were carried out with UA, UB, UC and Pt. For the uranyls, the anomalous difference were included. The E value was estimated from the RMS of lack of closure of isomorphous replacement at each round of refinement for each derivative, and E", lack of closure of anomalous contributions. Initial mean figure of merit is 0.80, and after three rounds of refinement, m=0.85 (for both CL & NT). Heavy atom parameters were listed in our previous report (3). Statistics of refinement are listed in Table 1. All the calculations were carried out on a GA mini computer with programs written in this laboratory.

Table 1 Statistics of Refinement

| Resolution of | | $R_{FHLE}$ | $R_k$ | | $\frac{RMS(F_H)}{RMS(E)}$ | | m | |
|---|---|---|---|---|---|---|---|---|
| Data | A | CL | CL | NT | CL | NT | CL | NT |
| UA | 3.5 | 0.48 | 0.078 | 0.091 | 1.78 | 1.72 | 0.85<br>(m would be 0.82 if the anomalous dispersion were incorrectly applied) | 0.85 |
| UB | 3.5 | 0.47 | 0.080 | 0.066 | 1.79 | 1.89 | | |
| UC | 3.5 | 0.45 | 0.082 | 0.096 | 1.92 | 1.89 | | |
| PT | 5.0 | - | 0.087 | 0.110 | 1.61 | 1.47 | | |

$R_{FHLE} = \sum_N |F_{HLE} - F_{H\ cal}| / \sum F_{HLE}$, RMS($F_H''$)/RMS($E''$) are between 0.7 and 0.9 for all the uranyls.

$R_k = \sum ||F_{PH}| - |F_p + F_H|| / \sum |F_{PH}|$

m = mean figure of merit

Total number of reflection in calculating best Fouriers are 3875 for CL and 3754 for NT. Unit weight was used throughout the least squares refinements.

Results and discussion: The electron density maps were calculated with coefficient mFexp($i\phi_B$) for both CL and NT. The grid is about 1 cm/A along the cell edges. In the maps, the average electron density of the main polypeptide chains is about 0.8 e/A$^3$ (0.4-2.0) and side groups, about 0.4 e/A$^3$ (0.2-1.0). Most of the large side groups are clearly seen. The main chain can be easily followed especially after the newly published results (1,2) reached us. Our maps are essentially in agreement with the 3A model (1) (Fig.1 CL). A break of continuity of the chain near the N-terminal is observed in both maps at the same location (Fig.2a-1 CL, 2a-2 NT). The ambiguity of electron density near the base of the active cleft where four chains met as described by the 3A model (1) is shown in Fig.2b (CL). The supposed active important Asp 31 and Asp 208 (in

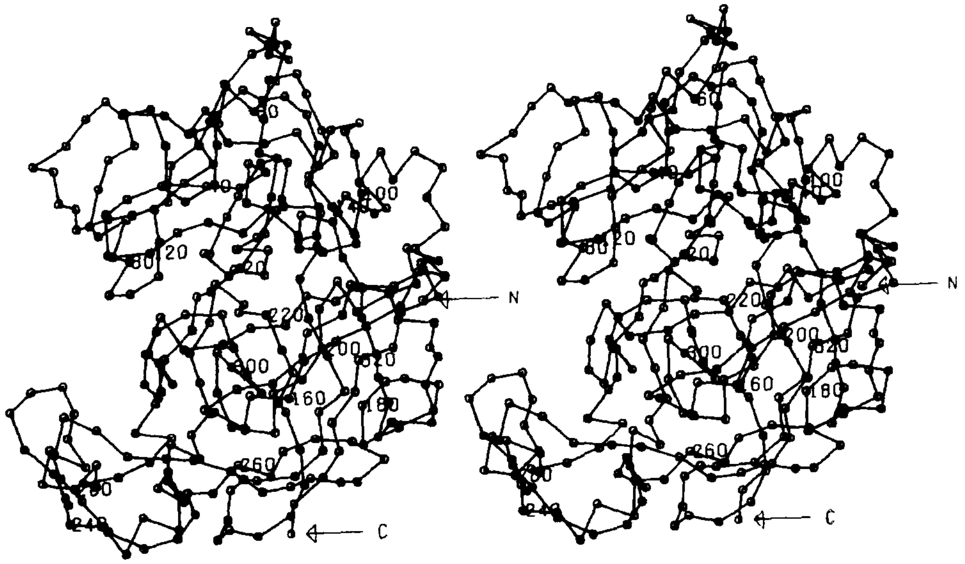

Fig.1 Stereo drawing of the main polypeptide chain in Endothia parasitica (CL). The circles represent α-carbon positions. Numbers are temporarily assigned pending further refinement at higher resolutions. N and C indicate amino and carboxy-terminals.

pepsin numbering (13); 33, and 223 in our numbering) are shown in Fig.2c. Both 33 and 223 have large side groups protruding into the solvent trough.

Both the N and C-terminals are very much like those of Rhizopus chinensis (1). The N-terminal for both CL and NT are shown in Fig.2a. The one on NT map (Fig.2a-2) should be interpreted as the 3A model (1). There are differences in $F_{CL}$ and $F_{NT}$($R=\sum||F_{CL}|-|F_{NT}||/\sum F_{CL}=0.1$) and in phase angles $<\Delta\alpha_B>=40^0$. The major features are almost identical except at the protruding parts in which the CL map gave more details and extended longer. All in all, the CL map could account for 326 residues and the NT map, about 315. The number of residues deduced from the CL map is in good agreement with the known amino acid composition of this protease, 328 (14). The other acid proteases are also composed of similar number of residues: pig pepsin 327 (13), Rizopus chinensis 324 (1) and penicillopepsin 323 (15). In the 3A model, only 305 residues were observed (1).

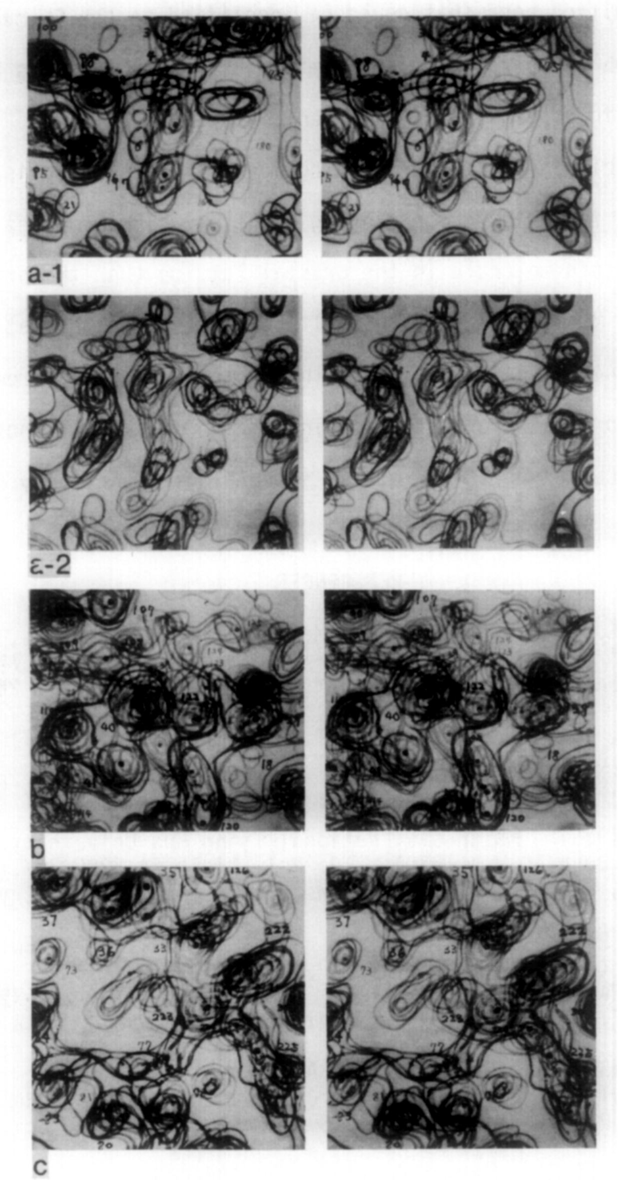

Fig.2 Stereoviews of electron density maps. Initial contours at 2.5 e/A$^3$ level. Increment at 2.5 e/A$^3$. Solid spots indicate approximate α-carbon positions, a-axis is along the horizontal direction. b-axis is along the normal direction. Interlayer distance is approximately 1A.

2a-1) Environment of N-terminal for CL. Eight layers (1.189-1.090). Residue 7 should be placed in the region of discontinuity.

2a-2) N-terminal for NT. Eight layers.

2b The region where four chains (29...32, 120...125, 38...41, and 85...90) meet. Ten layers (1.054-0.932).

2c Active groove. Solvent trough formed by chains 221...223, 33...37, and 73...79. 33 and 223 could be the two aspartic acid. Their protruding side groups are clearly seen. Six layers (0.919-0.851).

The only disulfide bond that could be identified in the CL map is formed between 258 and 285 (in our numbering) similar to Cys 250 and Cys 283 (in pig pepsin numbering) in pepsin (13) and in penicillopepsin (2).

The improved clarity of the protruding loops in the CL map is probably due to the network formed by cross-linking reaction which provides a better support of the protein·molecule in the crystal lattice than the original mother liquid (2.7 M ammonium sulfate). It is very fortunate that in spite of the radical change of environment for the protein molecule (from 2.7 M ammonium sulfate to pure water), the folding of the molecule is not observably affected, and the CL enzyme still retains its biological activity (3).

## REFERENCES

1. Subramanian, E., Swan, I.D.A., Liu, M., Davis, D.R., Jenkins, J.A. Tickle, I.J. & Blundell, T.L. (1977) Proc.Nat.Acad.Sci.USA 74, 556-559.
2. Hsu, I-N, Delbaere, L.T.J. & James, M.N.G. (1977) Nature, 266, 140-145.
3. Wong, C., Lee, T.J., Lee, T.Y., Lu, T.H. & Hung, C.S., (1977) Biochem. Biophys. Res. Commun.,
4. Hagemeyer, K., Fawwal, I. & Whitaker, J.R. (1968) J. Dairy Sci. 51, 1916-1922.
5. Cheng, N.T. (1975) MS thesis, National Tsing Hua University.
6. Moews, P.C. & Bunn, C.W. (1970) J. Mol. Biol. 54, 395-397.
7. North, A.C.T., Philips, D.C. & Mathews, F.S. (1966) Acta Cryst. A24, 351-354.
8. Singh, A.K. & Ramaseshan, S. (1966) Acta Cryst. 21, 279-280.
9. Mathews, B.W. (1965) Acta Cryst. 18, 745-749.
10. Kartha, G. & Parthasarathy, R. (1965) Acta Cryst. 18, 745-753.
11. Dickerson, R.E., Weinzierl, J.E. & Palmer, R.A. (1968) Acta Cryst. B24, 997-1003.
12. Blow, D.M. & Matthews, B.W. (1973) Acta Cryst. A29, 56-62.
13. Tang, J., Sepulveda, P., Marciniszyu Jr. J., Chen, K.C.S., Huang, W-Y., Tao, N., Liu, P. & Lanier, J.P. (1973) Proc. Nat. Acad. Sci. 70, 3437-3439.
14. Whitaker, J.R. (1971) Method Enzymol. 19, 436-445.
15. James, M.N.G., Hsu, I-N & Delbaere, L.T.J. (1977) Nature, 267, 808-813.