The EMBO Journal vol.3 no.6 pp.1311 – 1314, 1984

Three-dimensional structure of fungal proteinase K reveals similarity to bacterial subtilisin

A.Pähler[1,2], A.Banerjee[1], J.K.Dattagupta[1], T.Fujiwara[1], K.Lindner[1], G.P.Pal[1,2], D.Suck[1], G.Weber[1] and W.Saenger[1,2]*

[1]Max-Planck-Institut für experimentelle Medizin, Hermann-Rein-Str. 3, D-3400 Göttingen, and [2]Institut für Kristallographie, Freie Universität Berlin, Takustr. 6, D-1000 Berlin 33, FRG

*To whom reprint requests should be sent

Communicated by V.A.Erdmann

The three-dimensional structure of the fungal serine protease proteinase K has been determined at 3.3 Å resolution by single crystal X-ray diffraction analysis. The enzyme crystallizes in the tetragonal space group $P4_32_12$ with cell constants $a = b = 68.3$ Å, $c = 108.5$ Å. The asymmetric unit consists of one monomer of 27 000 daltons mol. wt., ~50% higher than the so far assumed value of 18 500 daltons. The main chain fold of proteinase K shows a high degree of tertiary homology with the corresponding bacterial subtilisin BPN'. Proteinase K is the second enzyme in this family of serine proteases to be studied by X-ray diffraction, thus confirming the existence of two unrelated families of serine proteases in pro- and eukaryotes.

Key words: proteinase K/X-ray structure/serine proteinase

Introduction

A highly active serine protease can be isolated from the culture filtrate of the fungus *Tritirachium album* Limber. It has been named proteinase K because of its capability to digest even native keratin (Ebeling *et al.*, 1974) and it remains active in the presence of low concentrations of SDS. This property has made proteinase K very useful for the isolation of native RNAs in high yields (Wiegers and Hilz, 1981), and most biochemical studies involving proteinase K concern mainly this application of the enzyme.

The serine proteases can be classified into the two distinct families, the chymotrypsin and the subtilisin families, which are unrelated to each other both in amino acid sequence and in tertiary structure. At the tertiary structure level, however, only subtilisin BPN' and subtilisin novo, which have identical amino acid sequences, are known as prototypes of their family, whereas the chymotrypsin family is represented by the structures of several different enzymes of pro- and eukaryotic origin. To investigate the structural relationship between proteinase K and the other serine proteases with known three-dimensional structure, an X-ray analysis of crystals of proteinase K was carried out. Based on a mol. wt. of 18 000 daltons, it was originally assumed that the asymmetric unit of the crystal contains a dimer (Dattagupta *et al.*, 1975). Since chymotrypsin-type serine proteases are in general of lower mol. wt. than the subtilisins, the mol. wt. of proteinase K suggested that this enzyme belongs to the chymotrypsin family. Contrary to this, however, was the finding that proteinase K and subtilisin Carlsberg produce similar cleavage patterns with oxidized insulin B chain as a substrate (Kraus *et al.*, 1976). Knowledge of the amino acid sequence around the active site serine, which is characteristic and identical within each family, would have answered this question unambiguously. However, at the time when our X-ray study was initiated, no details of the amino acid sequence were known.

Attempts to localize the suspected dimer in the electron density maps were unsuccessful and delayed the determination of the structure considerably. Finally it was decided to ignore the available biochemical data and rely only on the electron density map as a source of structural information. As a result, the electron density map could be interpreted by fitting to it one monomeric polypeptide chain. The determination of the amino acid sequence of proteinase K, which is presently under way (Jany and Pfleiderer, personal communication), confirmed our results and conclusions.

Results

The molecular envelope of proteinase K is an almost spherical ellipsoid with axes 38 Å, 43 Å and 38 Å. According to our chain tracing, the enzyme consists of 266 amino acid residues. The mol. wt. calculates to 27 000 daltons based: (i) on the overall dimensions of the molecule assuming a partial specific protein volume $\bar{v} = 0.74$ cm^3/g (Matthews, 1968); (ii) from the chain-length, assuming that the mol. wt. per peptide is ~100 daltons (Feldman, 1976) and (iii) from a comparison of the crystal unit cells of proteinase K and subtilisin BPN', when the solvent contents are appropriately accounted for (Wright *et al.*, 1969). This mol. wt. is ~50% higher than the previous value obtained by gel filtration (Ebeling *et al.*, 1974). An independent SDS gel electrophoresis, carried out in our laboratory after the structure was known, suggests a value of 30 000 ± 3000 daltons, in good agreement with the crystallographic data.

As can be seen from the two stereo drawings in Figure 1 and the artist drawing of the molecule in Figure 2, the wealth of secondary structure, containing several helices, is remarkable. This immediately rules out the possibility that proteinase K might belong to the chymotrypsin family of serine proteases. Even at a first glance the similarity of this structure with that of subtilisin BPN' becomes fairly obvious.

The core of the molecule is made up from a seven-stranded twisted parallel β-sheet and four long α-helices which run almost parallel to each other, the longest of them extending through the entire molecule. The direction of these helices is opposite to the mean direction of the sheet. A fifth helix, arranged at nearly right angles to the other four helices, can be seen in the C-terminal section of the molecule, and is followed by a somewhat questionable anti-parallel two-stranded β-sheet. Another anti-parallel sheet, again with two strands, precedes the central fourth helix.

Whereas the identification of the helices, except for the exact length, is quite straightforward, the assignment of the β-sheets is not so easy and should in principle be based on the corresponding hydrogen bonding patterns which at this resolution can of course not be given. Structure elements of the β-

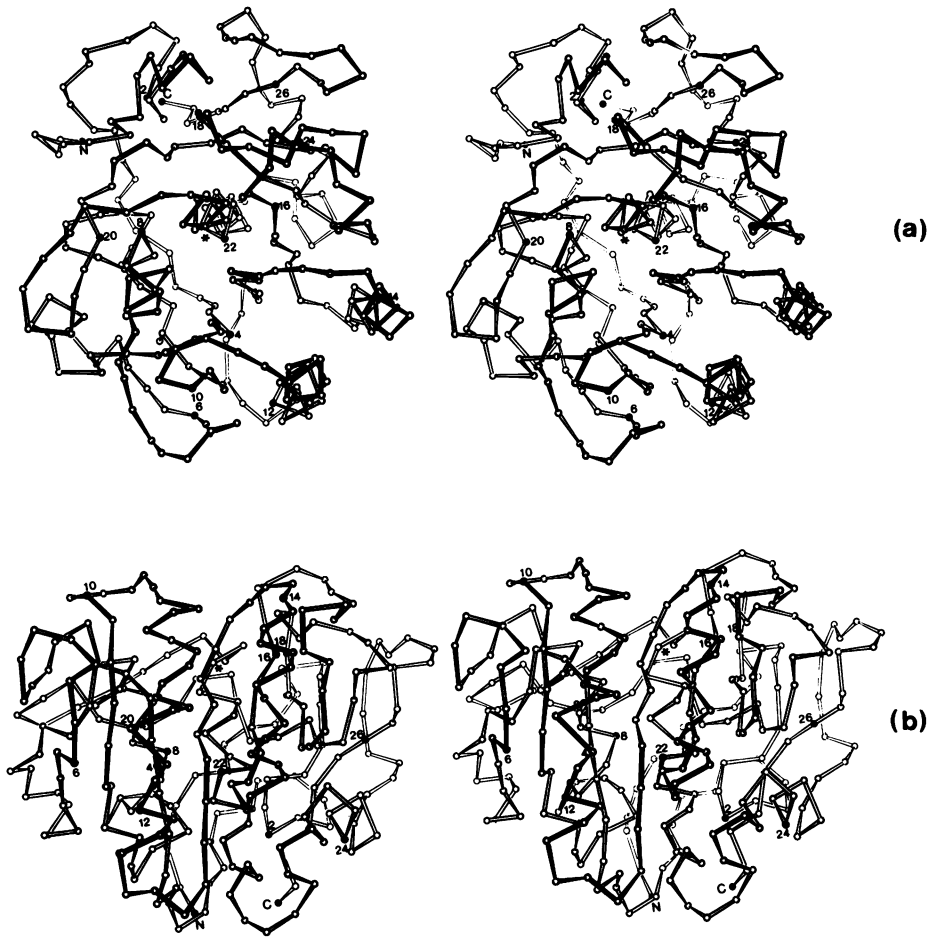

Fig. 1. **(a)** Stereo view of the α-carbon-chain of proteinase K along the crystallographic *c*-axis. Every 20th residue is marked black and residue numbers divided by 10 are indicated. The asterisk symbolizes the tentatively assigned active Ser213. **(b)** as **(a)** but viewed from the lower right.

pleated sheet type were derived from inspection of a list of C_α-C_α distances and their identification was checked against a diagonal distance plot and from comparison with subtilisin BPN'. The features indicative of β-sheet elements are clearly visible in the plot and a comparison of the diagonal distance plots (Remington and Matthews, 1980) of proteinase K and subtilisin BPN' (Figure 3) strongly suggests that the two enzymes are indeed very similar. Most of the deviations occur in the N- and C-terminal regions located close to each other at the surface of the molecule, whereas the central core is very well preserved.

A more quantitative comparison can be made by attempting to fit the coordinates of the C_α atoms. For this purpose the method of Kabsch (1976, 1978) was used and refined coordinates for subtilisin BPN' were taken from the Protein Data Bank (Bernstein *et al.*, 1977). A preliminary transformation matrix was obtained by identifying the amino acid residues forming the catalytic triad of proteinase K. The active site serine is followed by a residue with very high side chain density, assumed to be methionine, and histidine shows a well defined side chain. These two amino acids are located at the N-terminal ends of α-helices. Only aspartic acid is not clearly indicated and was tentatively placed at the C terminus of the first strand of the parallel β-sheet, where it is located in subtilisin BPN'. The triangles formed by the three active site residues were then rotated into a plane with a common center of gravity (Nishikawa and Ooi, 1974). If this is done, the root mean square (r.m.s.) deviation of the coordinates of the catalytic triads of the two proteins is 0.81 Å and the geometry is nearly preserved. Using the transformation matrix obtained in this way as a starting point, we attempted to find a subset of common coordinates by the criterion that corresponding atoms of the two proteins should be close together and the least possible number of amino acid insertions and deletions should be permitted. The subset thus chosen for the least squares fit, contained 199 C_α-carbon atoms, i.e., ~75% of the entire proteinase K molecule. After refinement of the transformation matrix against this subset of coordinates we compared the agreement of the secondary features of the core, comprising four of the five helices, the anti-parallel two-stranded β-sheet which is followed by the active site serine-containing α-helices, and the five prominent strands of the parallel sheet. This amounts to 90 residues or 34% of the molecule and yielded a r.m.s. difference of 1.4 Å between the coordinates of proteinase K and subtilisin BPN'. Considering the fact that the proteinase K coordinates were unrefined and

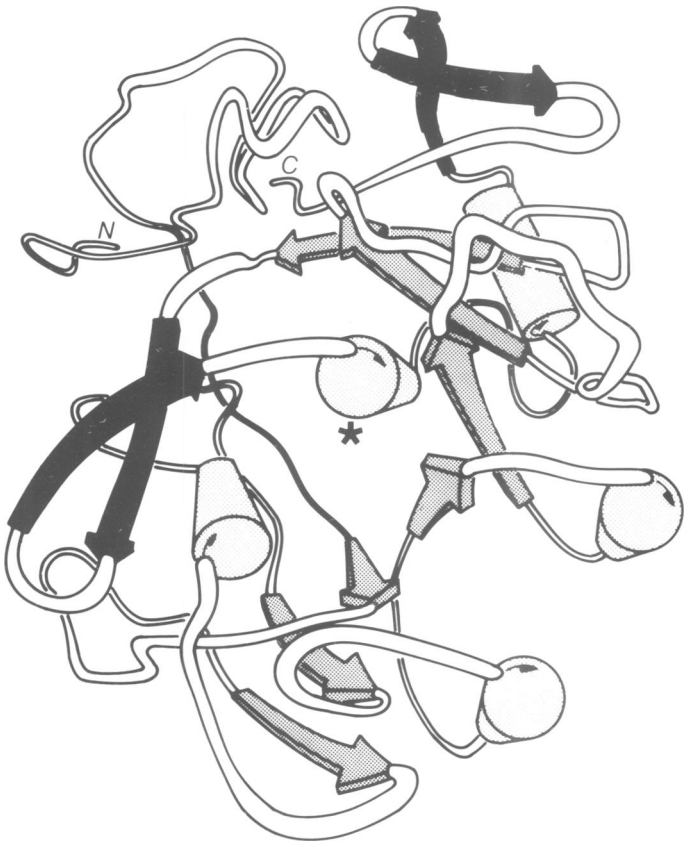

Fig. 2. Artist's drawing of the polypeptide chain of proteinase K in the same orientation as in Figure 1a. Helices are denoted by cylinders, β-sheet strands by arrows. An asterisk marks the presumable location of active Ser213 and C,N denote carboxy and amino termini.

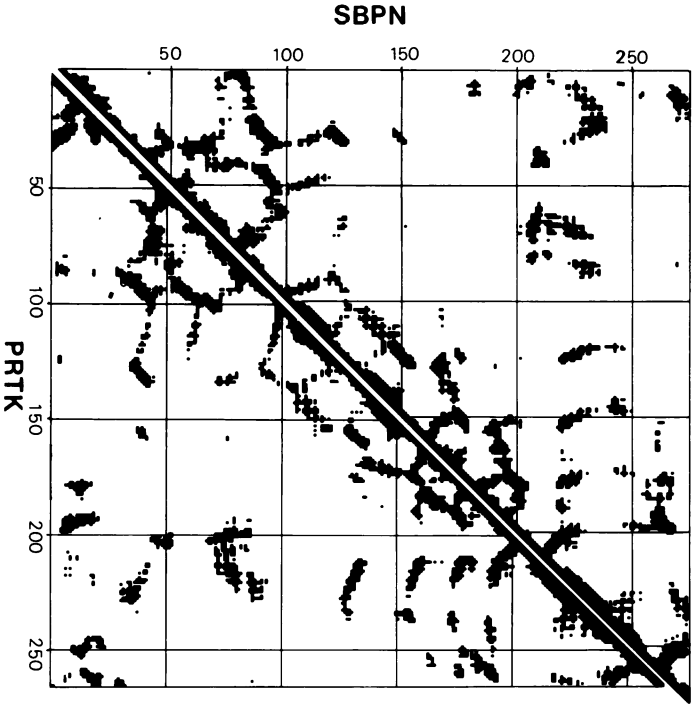

Fig. 3. Diagonal distance plots of proteinase K(PRTK) and subtilisin BPN′ (SBPN).

read off a mini-map, we can conclude that the arrangement of secondary structure in the two molecules is almost identical.

Discussion

From these results it must be concluded that the bacterial subtilisin BPN′ and the fungal proteinase K are evolutionarily related and originate from a common ancestor. This is interesting, because all other crystallographically studied serine proteases except subtilisin BPN′ have proved to belong to the chymotrypsin family (James, 1980). The amino acid sequence around the active site serine is identical within each family, a finding backed by a high degree of tertiary homology regardless of the biological origin of the respective protease. Our crystallographic study thus establishes firmly the existence of a second family of serine proteases. It speaks against the hypothesis that both families have evolved from a common ancestral enzyme (Hartley, 1970), and for the existence of two distinct lines of evolution. Comparable situations in the evolution of proteins and structural conservation in pro- and eukaryotes have been observed. Among them are the globins (Hendrickson and Love, 1971), the cytochromes (Almassy and Dickerson, 1978), the lysozymes (Grütter *et al.*, 1983), and the bacterial and fungal ribonucleases of the type RNase T1, whose structures have been determined recently (Hill *et al.*, 1983). Although differing widely in their amino acid sequence, the crystallographically studied ribonucleases display a high degree of homology as far as their secondary structures are concerned. This holds especially for the functional amino acid side chains of the active centers which are all preserved and in comparable orientations.

For a high resolution study, which should be the next step once the amino acid sequence is available, proteinase K offers good prospects. The crystals are fairly stable in the X-ray beam for ~200 h, they diffract well beyond 1.5 Å resolution with synchrotron radiation and can easily be obtained within one day. Without attempting to optimize their size, they quite often grow as large as 2.5–3 mm in their longest dimension and 1.0–1.5 mm in the other directions. Such crystals may henceforth be used for neutron diffraction studies, especially if they can be grown slightly larger. Data collection to a resolution of 1.5 Å is in progress. If phase expansion (Bricogne, 1974) starting from the now known model is successful, it will then be possible to verify the relationship of proteinase K and subtilisin BPN′ on the basis of a high resolution model of proteinase K, and make it a crystallographic prototype for this family of serine proteases.

Materials and methods

Proteinase K crystals up to 2 mm in size and diffracting well to at least 1.2 Å resolution, can be grown within 6–8 h by microdialysis (Dattagupta *et al.*, 1975) of a 10% protein solution against 1 M $NaNO_3$, 50 mM Tris, 10 mM $CaCl_2$, adjusted to pH 6.5, or by vapour diffusion against pure water. The material used was a highly purified preparation kindly supplied by Drs. Hennrich and Orth, Fa. E.Merck (Darmstadt). Crystals display space group $P4_32_12$ with cell constants $a = b = 68.3$ Å and $c = 108.5$ Å (Dattagupta *et al.*, 1975). Based on the number of amino acid residues per molecule found in this study, the mol. wt. of proteinase K was calculated to be 27 000 daltons. Therefore, the asymmetric unit contains one molecule and the solvent content is 48%.

X-ray diffraction data were collected using a STOE four-circle diffractometer, with Ni-filtered CuK_α radiation from a fine focus tube and employing ω-scan mode with stationary background measurement. Four data sets to

Table I. Heavy atom parameters utilized in phase determination

Atom	occ	x	y	z	B	R	P
PB 1	0.1063	0.802	0.515	0.499	5.88		
						0.380	2.36
Pb 2	0.1667	0.685	0.958	0.935	33.90		
Sm 1	0.2784	0.800	0.512	0.499	4.56		
						0.425	2.26
Sm 2	0.3997	0.690	0.946	0.936	65.60		
Hg 1	0.2404	0.723	0.758	0.698	8.31	0.415	2.52

occ: site occupancy on a relative scale; x, y, z: fractional coordinates; B: temperature factor in Å^2; R: weighted R-factor (Σ w$|F_{HLE}-F_H|^2/\Sigma$ w $F_{HLE}^2)^{1/2}$ using $1/\sigma^2$ weights; P: average phasing power $<F_H/E>$ where E denotes the total lack of closure error, F_H is the heavy atom structure factor amplitude.

2.8 Å resolution from native crystals were combined with a merging R factor of 4.4% in F and an average $<F/\sigma_F>$ of 78.2. Heavy atom derivatives were prepared with $Pb(NO_3)_2$, $HgCl_2$ and $SmCl_3$ by soaking native crystals in crystallization buffer containing the heavy atom salts. Crystals to be soaked in $SmCl_3$-containing solution were first dialysed extensively against 50% sodium citrate, then 0.3 M EDTA to remove Ca^{2+} ions present in the crystallization buffer (Martin and Richardson, 1979). One data set was collected for each of the derivatives, to 2.8 Å, 3.4 Å and 3.3 Å for the Pb, Hg and Sm derivatives, respectively, including measurement of Friedel pairs hkl and khl for each of them.

Heavy atom positions were determined by direct methods (Wilson, 1978) using MULTAN (Main *et al.*, 1980) with F_{HLE} estimates of the heavy atom scattering factor amplitudes (Matthews, 1966) and confirmed by Patterson syntheses. The heavy atom parameters were refined against all available F_{HLE} values by least squares minimization utilizing a modified ORFLS program (Busing *et al.*, 1977). Although Sm occupies the same positions as Pb (Table I), it was included in the phasing process because of its large anomalous dispersion. For the determination of protein phases the Blow and Crick (1959) procedure was employed. The calculation of phases was restricted to 3.3 Å resolution, since no good quality data for the Hg and Sm derivatives were available beyond this resolution and the Pb derivative is likely to be non-isomorphous with respect to the native crystals beyond 3.0 Å. For the whole data set of 4281 unique reflections, for which phases were calculated, the mean figure of merit was 0.846, and 3143 reflections had a figure of merit $m \geq 0.8$. This corresponds to a mean phase error of 32.3°. Details of the refinement and phasing results for the three heavy atom derivatives are given in Table I.

An electron density distribution was calculated with this phase set and contoured in z sections on a mini-map at a scale of 3.8 Å/cm in x and y. Interpretation of this map was fairly straightforward in most parts of the molecule. Only in those regions, which later proved to be the N and C termini of the enzyme, some difficulties arose and the results there should be treated with some caution.

Acknowledgements

We gratefully acknowledge the support of Drs. N.Hennrich and H.D.Orth (Fa. E. Merck, Darmstadt), through a continuous supply of highly purified proteinase K. All calculations were carried out in the computing center of the Gesellschaft für wissenschaftliche Datenverarbeitung in Göttingen. This study has been financially supported by Fonds der chemischen Industrie, by the award of an Alexander von Humboldt Fellowship to T.F. and by the Deutsche Forschungsgemeinschaft through Schwerpunktsprogramm and Sfb 9.

References

Almassy,R.J. and Dickerson,R.E. (1978) *Proc. Natl. Acad. Sci. USA,* **75**, 2674-2678.

Bernstein,F.C., Koetzle,T.F., Williams,G.J.B., Meyer,E.F.Jr., Brice,M.D., Rodgers,J.R. Kennard,O., Shimanouchi,T. and Tasumi,M. (1977) *J. Mol. Biol.,* **112**, 535-542.

Blow,D.M. and Crick,F.H.C. (1959) *Acta Crystallogr.,* **12**, 794-802.

Bricogne,G. (1974) *Acta Crystallog.,* **130**, 395-405.

Busing,W.R., Martin,K.O. and Levy,H.A. (1977) *ORFLS4,* Oak Ridge National Laboratory, Tennessee, USA.

Dattagupta,J.K., Fukiwara,T., Grishin,E.V., Lindner,K., Manor,P.C., Pieniazek,N.J., Saenger,W. and Suck,D. (1975) *J. Mol. Biol.,* **97**, 267-271.

Ebeling,W., Hennrich,N., Klochow,M., Metz,H., Orth,H.D. and Lang,H. (1974) *Eur. J. Biochem.,* **47**, 91-97.

Feldman,J.R. (1976) *AMSOM: Atlas of Molecular Structure on Microfiche,* Index Volume Tracor Jitco Inc., Rockville.

Grütter,M.G., Weaver,L.H. and Matthews,B.W. (1983) *Nature,* **303**, 828-831.

Hartley,B.S. (1970) *Phil. Trans. R. Soc. London;* **B257**, 77-87.

Hendrickson,W.H. and Love,W.E. (1971) *Nature,* **232**, 197-203.

Hill,C., Dodson,G., Heinemann,U., Saenger,W., Mitsui,Y., Nakamura,K., Borisov,S., Tischenko,G., Polyakov,K. and Pavlovsky,S. (1983) *Trends Biochem. Sci.,* **8**, 364-369.

James,M.N.G. (1980) *J. Biochem.,* **58**, 251-271.

Kabsch,W. (1976) *Acta Crystallogr.,* **A32**, 422-423.

Kabsch,W. (1978) *Acta Crystallogr.,* **A34**, 827-828.

Kraus,E., Kiltz,H.H. and Femfert,U. (1976) *Hoppe-Seyler's Z. Physiol. Chem.,* **357**, 223-237.

Main,P., Fiske,S.J., Hull,S.E., Lessinger,L., Germain,G., Declercq,J.P. and Woolfson,M.M. (1980) *MULTAN 80,* University of York, England, and Louvain, Belgium.

Martin,R.B. and Richardson,F.S. (1979) *Qt. Rev. Biophys.,* **12**, 181-209.

Matthews,B.W. (1966) *Acta Crystallogr.,* **20**, 230-239.

Matthews,B.W. (1968) *J. Mol. Biol.,* **33**, 491-497.

Nishikawa,K. and Ooi,T. (1974) *J. Theor. Biol.,* **43**, 351-374.

Remington,S.J. and Matthews,B.W. (1980) *J. Mol. Biol.,* **149**, 77-99.

Wiegers,U. and Hilz,H. (1981) *Biochem. Biophys. Res. Commun.,* **44**, 513-519.

Wilson,K.S. (1978) *Acta Crystallogr.,* **B34**, 1599-1608.

Wright,L.S., Alden,R.A. and Kraut,J. (1969) *Nature,* **221**, 235-242.

Received on 22 February 1984