Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

1.7 Å X-ray structure of space-grown collagenase crystals

I. Broutin-L'Hermite,* M. Ries-Kautt and A. Ducruix

Laboratoire d'Enzymologie et de Biochimie Structurales, Bâtiment 34, CNRS, 91198 Gif sur Yvette CEDEX, France

Correspondence e-mail: broutin@lebs.cnrs-gif.fr

Received 10 December 1999
Accepted 22 December 1999

PDB Reference: collagenase, 2hlc.

Collagenases, divided into metallocollagenases and serine collagenases, are the only proteases that cleave the triple helix of collagen under physiological conditions. In the present work, the serine protease collagenase purified from *Hypoderma lineatum* larvae is studied. From crystals grown in the International Microgravity Laboratory (IML2), a data set was collected at 1.7 Å using synchrotron radiation. Although the resolution is not very different, the signal-to-noise ratio and the quality of the electron density are much improved. Alternate conformations were revealed for several residues, in particular Tyr99, suggesting a gate mechanism of recognition.

1. Introduction

Collagenases are the only enzymes able to cleave native collagen under physiological conditions. Two different classes of collagenases exist: the metallocollagenases, which are found mostly in vertebrates, and the serine collagenases, which belong to the invertebrate family. The latter enzymes are involved in digestive rather than morphogenic functions. Several structures of collagenases from the two families have been solved over the last five years (Li *et al.*, 1995; Gomis-Ruth *et al.*, 1996; Broutin *et al.*, 1996; Betz *et al.*, 1997; Perona *et al.*, 1997; Brandstetter *et al.*, 1998; Moy *et al.*, 1999), but no similarities could be found, metallocollagenase structures being mostly helical and serine collagenase being mostly constituted of β-sheets.

The collagenase used in this study belongs to the chymotrypsin family (Boulard & Garrone, 1978). It is extracted from the digestive tract of the larvae *H. lineatum*, a cattle parasite responsible for hypodermosis. *H. lineatum* collagenase (HLC) is composed of 230 amino-acid residues, corresponding to a mass of 25 209 Da. It cleaves type III collagen in its helical part in only one region, producing the two fragments (3/4 and 1/4 length) characteristic of the cleavage of collagen by vertebrate collagenases (Lecroisey & Keil, 1985), but so far the mechanism by which collagenase cleaves collagen is unknown. The structure of the enzyme, initially determined at 1.8 Å resolution (Broutin *et al.*, 1996), was refined using a new data set from a space-grown crystal. This data set had a signal-to-noise ratio which was clearly better than those obtained from ground-grown crystals (Fig. 1*a*).

2. Materials and methods

2.1. Protein crystallization and X-ray diffraction data

Collagenase was purified from *H. lineatum* larvae according to a previously described procedure (Lecroisey *et al.*, 1979). Crystals were grown under reduced gravity during the IML2 mission of the US Space Shuttle, which flew for 14.5 d in July 1994 (Broutin *et al.*, 1997), using the APCF apparatus (Bosch *et al.*, 1992). One crystal (1.2 × 0.8 × 0.8 mm) was mounted in a glass capillary of 2 mm diameter. Measurements were made 5 d after landing, at 279 K on the wiggler beamline W32 (Fourme *et al.*, 1992) at LURE (Orsay, France) using a wavelength of 0.901 Å and a 300 mm diameter MAR Research image-plate detector. Complete diffraction data were collected from a single crystal. The total oscillation range was 55°, recorded in steps of 1° with an exposure time of 180 s per image. The space group is $I422$, with two molecules in the asymmetric unit. The unit-cell parameters were refined to $a = 111.3$, $c = 165.7$ Å, which is isomorphous to the previously refined 1.8 Å structure (PDB code 1hyl). The data set was integrated using *MOSFLM* software (Leslie, 1987) and the *CCP*4 package (Collaborative Computational Project, Number 4, 1994). It contains 54 934 unique reflections from 19 to 1.7 Å resolution. The completeness and percentage of data with $I/\sigma > 3$ are 96.3 and 77.6%, respectively, over the total resolution range and 99.6 and 40%, respectively, in the last shell (1.7–1.74 Å). The global $R_{\rm sym}$ is 4.7% at 1.7 Å resolution for a multiplicity of 4.4 and the mean I/σ value is 11.4.

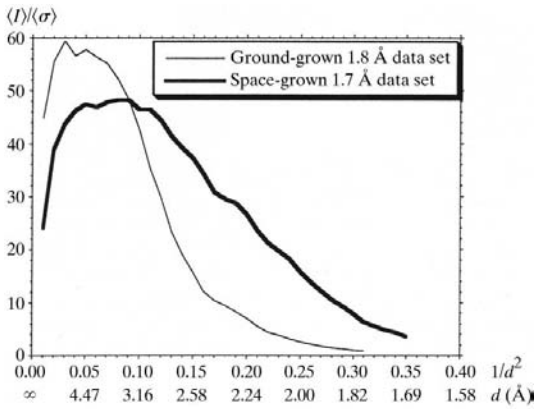

Figure 1
Two examples of quality improvement in the 1.7 Å HLC data set compared with the previous data set. (*a*) Quality improvement in reciprocal space: I/σ as a function of $1/d^2$; for a detailed comparison, see Broutin *et al.* (1997). (*b*) Example of the electron-density quality of HLC at 1.7 Å resolution.

2.2. Refinement

Refinement of the structure was performed in the resolution range 8–1.7 Å without any σ-cutoff, using *X-PLOR* version 3.851 (Brünger, 1992*b*). 10% of the data set in the 19–1.7 Å resolution range were omitted from the refinement and were used to calculate the $R_{\rm free}$ value (Brünger, 1992*a*). In order for the refinement not to be biased, all previously determined water molecules (PDB code 1hyl) except the 72 having the lowest temperature factors were omitted from the starting model.

During the first cycle, a slow-cooling procedure was executed, starting at a temperature of 3000 K. Subsequent cycles were limited to energy-minimization calculation and individual temperature-factor refinement. Between each cycle, $2F_o - F_c$ and $Fo - F_c$ maps were calculated between 19 and 1.7 Å and visualized using the graphic program *O* (Jones, 1978). The values of the *R* factor and $R_{\rm free}$ were 22.9 and 28.3%, respectively, after the first step and 19.5 and 24.5% after the last one. The mean temperature factors are 22 and 26 Å^2 for the two molecules (compatible with the Wilson slope of 23 Å^2) and 42 Å^2 for waters. The accuracy of the coordinates as determined by the program *SIGMAA* (Read, 1986) is 0.2 Å. Deviations from ideal geometry are 0.009 Å for bond lengths, 1.53° for bond angles, 26.82° for dihedrals and 1.27° for impropers.

3. Results and discussion

The structure of HLC was previously solved and refined at 1.8 Å resolution (PDB code 1hyl) using *X-PLOR* version 2.1 (Brünger, 1990) with no $R_{\rm free}$ factor and with different geometry parameters and topology files to the current version. The quality of the new data set with a three times better signal-to-noise ratio (Broutin *et al.*, 1997), as well as a completeness of 100% at 1.8 Å (the completeness of the previous data set was 74% on all data and only 51% in the last shell), prompted us to start a new refinement procedure. The electron density calculated from the new data set at 1.7 Å shows an amazing improvement in quality compared with the 1.8 Å map. An example of the map quality is given in Fig. 1(*b*). This improvement not only reflects the minor resolution shift, but also the fact that the number of individual reflections increased from 45 152 to 54 934, which underlines the dramatic effect of completeness on map calculation. Another indication of map quality is given by the fact that ten alternate positions were visualized and refined (four in molecule *A* and six in molecule *B*), only two of which were common to both molecules. Among those alternate positions, four residues (39 and 99 in molecule *A*, and 99 and 143 in molecule *B*) are quite remarkable. Arg*A*39[1] and Gln*B*143 are located at the interface between molecules *A* and *B*. The electron density shows that the extremities of one glutamine and of one arginine occupy the same position, thus leading to a mutually exclusive situation. Among the three different pairing possibilities, only one implies a direct hydrogen bond.

The other residue presenting an interesting alternate position is Tyr99 (Fig. 1*b*), which is located on one side of the entrance of the active site. In the previous refinement, its location prevented the entrance of any modelled peptide inhibitor because of steric hindrance. Binding of a peptide would therefore imply a rotation of Tyr99 towards the solvent, unlocking the active site. In the current refinement, the alternate position, present in both molecules, is in the open position, supporting the hypothesis that this residue may act as a gate controlling access to the active site.

This hypothesis is reinforced by the comparison of the HLC dimer structure with the X-ray structure of human leukocyte elastase (HLE; An-Zhi *et al.*, 1988) complexed with MeOSuc-A^{15}APV^{18}-CH$_2$Cl (PDB code 1hne; An-Zhi *et al.*, 1988). This inhibitor has been proven to inhibit HLC (data not shown). The model from the direct superposition of HLE and HLC (Fig. 2) indicates that the position of the inhibitor would be compatible with the formation of two covalent bonds between MeOSuc-AAPV-CH$_2$Cl and HLC in the real complex. There are also two potential hydrogen bonds between Ala*I*16 and Val*A*216. The main hindrance between the inhibitor and HLC arises from a steric clash between Pro*I*17 and Tyr*A*99. This clash is relieved by the alternate position of Tyr*A*99 which opens the access to the active site, supporting a gating mechanism.

On the opposite side of the active-site entrance, in a quite hydrophobic region, is the phenylalanine *A*192 [only present in HLC, HLE and PR3 (Fujinaga *et al.*, 1996; PDB code 1fuj) of all of the serine proteases] which seems to act as a pair of jaws with Tyr*A*99, limiting the access to the active site.

3.1. Importance of β-hairpin 33–41 (L^{33}-DITL^{37}-Q^{37A}-D^{37B}-Q^{37C}-R^{38}-RVW^{41}; chymotrypsin numbering)

All the members of the serine-protease family present the same secondary structure formed by two β-barrel domains. However, the length and tertiary structure of loops 33–41 are quite different. In HLC, this region forms a twisted β-hairpin rigidified by

[1] Hereafter, numbers are preceded by *A* for molecule *A* of the asymmetric unit and by *B* for molecule *B*.

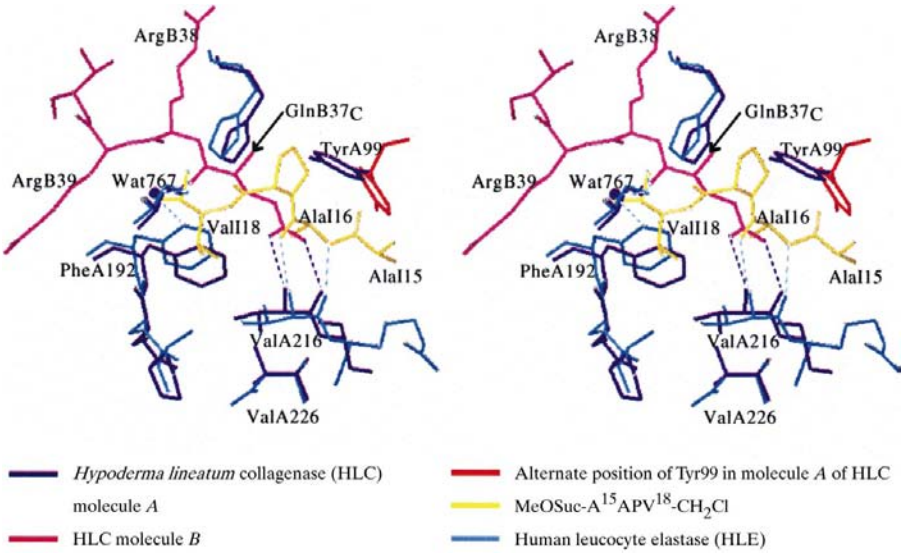

Figure 2
Superposition of the HLE–AAPV-CH_2Cl complex with HLC structure: a view of the active-site region.

internal hydrogen bonds. When compared with its equivalent loops in nine other serine proteinases, it may be asked whether the position of this region could be used as a cover to prevent access to the entrance of the active site.

This β-hairpin is at the interface between the two molecules, cross-interacting with the substrate-binding site of the other molecule *via* water molecules acting as a relay. It also forms one hydrogen bond with an alternate position of residue 99 implicated in gating the active-site entrance. Most of the contacts between molecules *A* and *B* in this region involve the three residues Asp37B, Gln37C and Arg38, which form a kind of tripod sitting on the active site of the second molecule.

The interlocking of the two molecules of the asymmetric unit through the 33–41 β-hairpin led us to the hypothesis that this may result in a self-inhibition. This would explain why HLC does not suffer from autolysis at optimum pH, contrary to most serine proteases.

4. Conclusions

Proteinase inhibitors still receive a great deal of attention because of their biological implications in many human and animal diseases. The clear structural homology between HLE and HLC allowed us to model enzyme-inhibitor specific features implicated in the activity of the two enzymes underlying the importance of three residues (Tyr99, Phe192 and Val216) in the process of substrate or inhibitor selection. Tyr99 is involved in a gate mechanism at the entry of the binding site. Phe192 and Val216, which are common to HLE and HLC, delineate a shallow hydrophobic binding cleft prohibiting the entry of large residues. Val216 is seemingly responsible for the formation of crucial hydrogen bonds between inhibitors and the enzyme.

The superposition of the two structures led us to the possibility of a self-inhibition mechanism in HLC involving the β-hairpin 33–41 (LDITLQDQRRVW).

We have described the extremity as well as both sides of the binding site close to the catalytic residues and this should help in the design of peptide inhibitors with high affinity.

Collagenase has been used as a test molecule to calibrate various detectors at synchrotron facilities and in laboratories. We therefore have a good knowledge of its diffraction pattern, with strong indications that although the improvement in resolution for space-grown crystals was minor, the signal-to-noise ratio was considerably improved. Of the many possibilities for explaining why the space-grown crystal yielded a better signal-to-noise ratio, the mosaic spread, crystal size and quality of the beam and detector have all been considered (Broutin *et al.*, 1997); however, no clear correlation could be firmly established.

IB was supported by the Centre National d'Etudes Spatiales (CNES). We acknowledge the European Space Agency (ESA) for providing crystallization reactors. This work was financially supported by the Centre Nationale de la Recherche Scientifique (CNRS). We are grateful to F. Pochon for preliminary activity tests of HLC on serine protease substrates. We would like to thank C. Boulard for providing larvae and A. Lecroisey for useful discussion. We are indebted to R. Fourme and J.-P. Benoit for the development of the W32 station (LURE, Orsay). We thank Glaxo for providing the unreleased metallocollagenase coordinates.

References

An-Zhi, W., Mayr, I. & Bode, W. (1988). *FEBS Lett.* **234**, 367–373.
Betz, M., Huxley, P., Davies, S. J., Mushtaq, Y., Pieper, M., Tschesche, T., Bode, W. & Gomis-Ruth, F. X. (1997). *Eur. J. Biochem.* **247**, 356–363.
Bosch, R., Lautenschlager, L., Potthast, L. & Stapelmann, J. (1992). *J. Cryst. Growth*, **122**, 310–316.
Boulard, C. & Garrone, R. (1978). *Comput. Biochem. Physiol.* **59**B, 251–255.
Brandstetter, H., Engh, R. A., Von Roedern, E. G., Moroder, L., Huber, R., Bode, W. & Grams, F. (1998). *Protein Sci.* **7**, 1303–1309.
Broutin, I., Arnoux, B., Riche, C., Lecroisey, A., Keil, B., Pascard, C. & Ducruix, A. (1996). *Acta Cryst.* D**52**, 380–392.
Broutin, I., Riès-Kautt, M. & Ducruix, A. (1997). *J. Cryst. Growth*, **181**, 97–108.
Brünger, A. T. (1990). *X-PLOR Manual, Version* 2.1. Yale University, New Haven, Connecticut.
Brünger, A. T. (1992*a*). *Nature (London)*, **355**, 472–475.
Brünger, A. T. (1992*b*). *X-PLOR Version* 3.851. Yale University, New Haven, Connecticut.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Fourme, R., Dhez, P., Benoit, J. P., Kahn, R., Dubuisson, J. M., Besson, P. & Frouin, J. (1992). *Rev. Sci. Instrum.* **63**, 63.
Fujinaga, M., Chernaia, M. M., Halenbeck, R., Koths, K. & James, M. N. G. (1996). *J. Mol. Biol.* **261**, 267–278.
Gomis-Ruth, F. X., Gohlke, U., Betz, M., Knauper, V., Murphy, G., Lopez-Otin, C. & Bode, W. (1996). *J. Mol. Biol.* **264**, 556–566.
Jones, T. A. (1978). *J. Appl. Cryst.* **11**, 268–272.
Lecroisey, A., Boulard, C. & Keil, B. (1979). *Eur. J. Biochem.* **101**, 385–393.
Lecroisey, A. & Keil, B. (1985). *Eur. J. Biochem.* **152**, 123–130.
Leslie, A. G. W. (1987). *Proceedings of the CCP4 Study Weekend. Computational Aspects of Protein Crystal Data Analysis*, pp. 39–50. Warrington: Daresbury Laboratory.
Li, J., Brick, P., O'Hare, M. C., Skarzynski, T., Lloyd, L. F., Curry, V. A., Clark, I. M., Bigg, H. F., Hazleman, B. L., Cawston, T. E. & Blow, D. M. (1995). *Structure*, **3**, 541–549.
Moy, F. J., Chanda, P. K., Chen, J. M., Cosmi, S., Edris, W., Skotnicki, J. S., Wilhelm, J. & Powers, R. (1999). *Biochemistry*, **38**, 7085–7096.
Perona, J. J., Tsu, C. A., Craik, C. S. & Fletterick, R. J. (1997). *Biochemistry*, **36**, 5381–5392.
Read, R. J. (1986). *Acta Cryst.* A**42**, 140–149.