Acta Cryst. (1996). D**52**, 380–392

1.8 Å Structure of *Hypoderma lineatum* Collagenase: a Member of the Serine Proteinase Family

ISABELLE BROUTIN,[a]*‡ BERNADETTE ARNOUX,[a]‡ CLAUDE RICHE,[a] ANNE LECROISEY,[b] BORIVOJ KEIL,[a]† CLAUDINE PASCARD[a] AND ARNAUD DUCRUIX[a]‡

[a]*Institut de Chimie des Substances Naturelles, CNRS 91198 Gif sur Yvette CEDEX, France, and* [b]*Institut Pasteur, 28 rue du Dr Roux, 75724 Paris CEDEX 15, France. E-mail: ducruix@cygne.lbs.cnrs-gif.fr*

(Received 17 April 1995; accepted 25 August 1995)

Abstract

Collagenase from the fly larvae *Hypoderma lineatum* cleaves triple-helical collagen in a single region. It was crystallized at neutral pH in the absence of inhibitor and 1.8 Å data were collected using synchrotron radiation and a Mark II prototype detector. The structure was solved by combining multiple isomorphous replacement methods and rotation translation function in real space. Refinement between 7 and 1.8 Å using the program *X-PLOR* led to a final *R* factor of 16.9%. The overall fold is similar to that of other trypsin-like enzymes but the structure differs mainly by the presence of a β-sheet at position 31–44. The two embedded molecules of the asymmetric unit are related by a pseudo twofold axis. The β-sheet 31–44 of one molecule is involved in hydrogen bonds with binding-pocket residues of the other molecule. It thus completely prevents access to the active site. The specificity of this enzyme probably results from the position of Phe192 and Tyr99 at the entrance of the active site.

1. Introduction

Collagenases are enzymes which cleave triple helical collagen under physiological conditions of pH, temperature and ionic strength (Mandl, 1961). In the collagenase family, as for thermolysin/trypsin enzymes, two different enzymatic mechanisms exist for the hydrolysis of a peptide bond. One requires a zinc to polarize the scissile peptide bond near the catalytic histidine, the other uses the well known catalytic triad Asp, His, Ser to cleave the peptide bond.

Most collagenases belong to the Zn-metalloenzyme family and several X-ray structures of catalytic domains of collagenase belonging to the matrix metalloproteinase family were recently solved (Bode *et al.*, 1994; Borkakoti *et al.*, 1994; Grams *et al.*, 1995; Li *et al.*, 1995; Lovejoy *et al.*, 1994; Spurlino *et al.*, 1994; Stams *et al.*, 1994). Collagenase from fly larvae *Hypoderma lineatum* is a member of the collagenolytic enzymes related to the trypsin family (Lecroisey, Boulard & Keil, 1979). This group consists of collagenases with digestive rather than morphogenic functions.

In this paper, we describe the three-dimensional structure of a collagenase purified from the larvae *H. lineatum* (hereafter referred to as HLC). Its collagenolytic activity in the midgut of the first instar migrating larvae from *H. lineatum* was first demonstrated by Boulard (1970). These larvae are endoparasites of cattle. Because their midgut is closed at one extremity, it acts as a reservoir for degradation products of the connective tissue of the host and for re-absorbed collagenase. As a consequence, large amounts of collagenase could be obtained.

The collagenase purified from *H. lineatum* larvae is a monomeric enzyme of molecular weight 25 223 Da with 230 amino-acid residues. It is stoichiometrically inhibited by di-isopropylfluorophosphate (Lecroisey *et al.*, 1979) as are all serine proteinases. Its amino-acid sequence was determined by chemical means (Lecroisey, Gilles, De Wolf & Keil, 1987) and from the cDNA sequence (Moiré, Bigot, Periquet & Boulard, 1994). HLC degrades type I and III collagen in the form of reconstituted rat skin fibrils, but it is not as accurate as human collagenase which cleaves collagen only between residues 775 and 776. In fact several cuts are observed between residues 764 and 801 (Lecroisey *et al.*, 1979). It also hydrolyses casein and the B chain of insulin.

2. Materials and methods

2.1. *Protein purification and crystallization*

The procedure for the preparation and purification of collagenase was described elsewhere (Lecroisey *et al.*, 1979). Crystals were obtained either by dialysis or vapor-diffusion techniques (Ries-Kautt & Ducruix, 1992). As the enzyme does not suffer autolysis, crystallization could take place at neutral pH close to the optimum pH level for activity (8–8.5). The best crystals were obtained with a 20 mg ml^{-1} solution of the HLC in 50 m*M* Tris pH 7.4 containing 60 m*M* NaCl and 1.23 *M* ammonium sulfate (Ducruix, Arnoux, Pascard, Lecroisey & Keil, 1981). The temperature of crystallization was kept con-

‡ Present address: Laboratoire de Biologie Structurale, Bâtiment 34, CNRS 91198 Gif sur Yvette CEDEX, France.
† Deceased (1994).

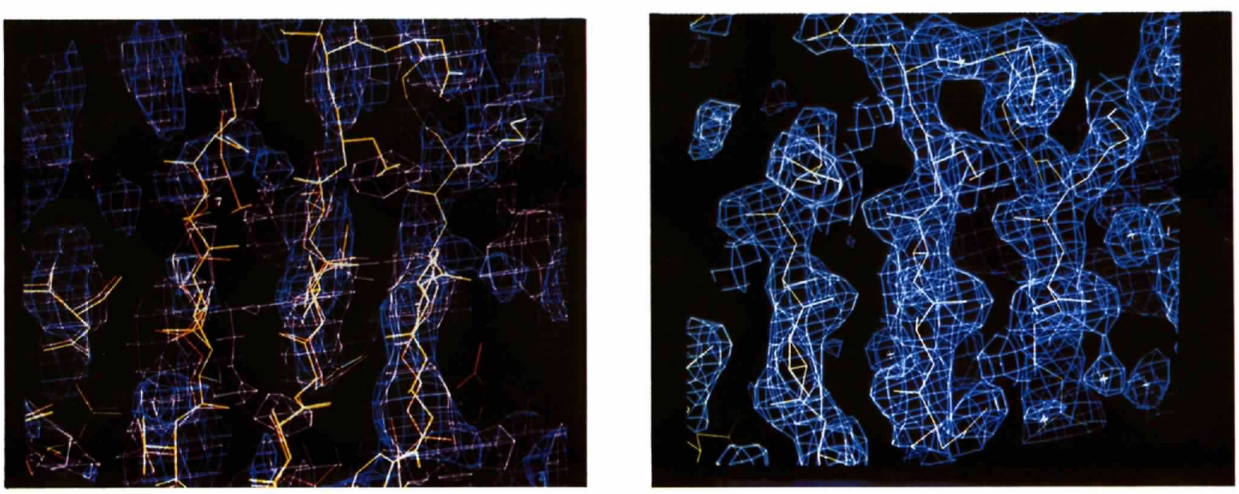
```
          16  20      25      30      35  37A 37B 37C 40      45      50      55      59A 60      65              80
HLC       IINGYEA YTG LFPYQAGLD ITLQDQRRV WCGGSLID NKWILTAAHCV HDA-------V-SVVVYL GSAV Q--YEG--E
HLE       IVGGRRA RPH AWPFMVSLQ LR----GGH FCGATLIA PNFVMSAAHCV ANV------NVRAVRVVL GAHN LSRREPT-R
ELA       VVGGTEA QRN SWPSQISLQ YRSGSSWAH TCGGTLIR QNWVMTAAHCV DRE-------L-TFRVVV GEHN LNQNNG-TE
CHT       IVNGEEA VPG SWPWQVSLQ DKT---GFH FCGGSLIN ENWVVTAAHCG --V-------TTSDVVVA GEFD QGSSSE-KI
TRP       IVGGYTC GAN TVPYQVSLN S-----GYH FCGGSLIN SQWVVSAAHCY --K-------S-GIQVRL GEDN INVVEG-NE
KAL       IIGGREC EKN SHPWQVAIY HY----SSF QCGGVLVN PKWVLTAAHCK --N-------D-NYEVWL GRHN LFENEN-TA
RMC       IIGGVES IPH SRPYMAHLD IVTEKGLRV ICGGFLIS RQFVLTAAHCK --G-------R-EITVIL GAHD VRKRES-TQ
SGT       VVGGTRA AQG EFPFMVRLS --------M GCGGALYA QDIVLTAAHCV SGSG-----NNTSITATG GVVD LQSGA---A
THR       IVEGSDA EIG MSPWQVMLF RKSP--QEL LCGASLIS DRWVLTAAHCL LYPPWDKNFTENDLLVRI GKHS RTRYERNIE

          85      90      95          100     105     109 111 115     120     125A 125B 130             140
HLC       AVVNSE RIISH S MFN P----DTY--L NDVALIKI -P-HVEYTDNIQ PIRLP SGEE-LNNKF-EN IWATVSGW GQS
HLE       QVFAVQ RIFED - GYD P----VNL--L NDIVILQL -NGSATINANVQ VAQLP A---QGRRLG-NG VQCLAMGW GLL
ELA       QYVGVQ KIVVH P YWN T----DDVAAG YDIALLRL AQ-SVTLNSYVQ LGVLP R----AGTILANN SPCYITGW GLT
CHT       QKLKIA KVFKN S KYN S----LTI--N NDITLLKL -STAASFSQTVS AVCLP S----ASDDFAAG TTCVTTGW GLT
TRP       QFISAS KSIVH P SYN S----NTL--N NDIMLIKL -KSAASLNSRVA SISLP T------SCASAG TQCLISGW GNT
KAL       QFFGVT ADFPH P GFN LSADGKDY--S HDLMLLRL QS-PAKITDAVK VLELP T------QEPELG STCEASGW GSI
RMC       QKIKVE KQIIH E SYN S----VPN--L HDIMLLKL -EKKVELTPAVN VVPLP S----PSDFIHPG AMCWAAGW GKT
SGT       VKVRST KVLQA P GYN G------T--G KDWALIKL -A-QPIN---QP TLKIA TT----TAYNQ-- GTFTVAGW GAN
THR       KISMLE KIYIH P RYN W----RENL-D RDIALMKL KK-PVAFSDYIH PVCLP DRETA-ASLLQAG YKGRVTGW GNL

          145 147         152     155     160     165     170 170A    175     180     185 186A    190     195     203 206
HLC       NTD---------TV ILQYTYNLVIDNDRC AQ-EYPPGIIVES TICGD -TSD---G KSPCFGDSGGPFV LSD----KNL
HLE       GR------NRGIAS VLQELNVTVVTSL-C ----------RRS NVCTL VRGR---Q AGVCFGDSGSPLV CN------GL
ELA       RT------NGQLAQ TLQQAYLPTVDYAIC SSSSYWGSTVKNS MVCAG -GD--G-V RSGCQGDSGGPLH CLVN--GQYA
CHT       RY-------ANTPD RLQQASLPLLSNTNC KK--YWGTKIKDA MICAG -AS----G VSSCMGDSGGPLV CKKN--GAWT
TRP       KSS-----GTSYPD VLKCLKAPILSDSSC KS--AYPGQITSN MFCAG -YLE-G-G KDSCQGDSGGPVV CS------GK
KAL       EPGPD---DFEFPD EIQCVQLTLLQNTFC AD--AHPDKVTES MLCAG -YLP-G-G KDTCMGDSGGPLI CN------GM
RMC       GV------RDPTSY TLREVELRIMDEKAC VD-YRY--YEYKF QVCVG SPTT---L RAAFMGDSGGPLL CA------GV
SGT       RE------GGSQQR YLLKANVPFVSDAAC RS-AYGNELVANE EICAG -YPDTG-G VDTCQGDSGGPMF RKDN-ADEWI
THR       KETWTANVGKGQPS VLQVVNLPIVERPVC KD--STRIRITDN MFCAG YKPDEGKR GDACEGDSGGPFV MKSPFNNRWY

          210     215     217A 220     225     230     235     240     245
HLC       LIGVVSFV SGAGCESG-KPV GFSRV TSYMDWIQQNTG IKF-
HLE       IHGIASFV R-GGCASGLYPD AFAPV AQFVNWIDSIIQ ----
ELA       VHGVTSFV SRLGCNVTRKPT VFTRV SAYISWINNVIA SN--
CHT       LVGIVSWG S-STC-STSTPG VYARV TALVNWVQQTLA AN--
TRP       LQGIVSWG S--GCAQKNKPG VYTKV CNYVSWIKQTIA SN--
KAL       WQGITSWG HT-PCGSANKPS IYTKL IFYLDWIDDTIT ENP-
RMC       AHGIVSYG HPDA----KPPA IFTRV STYVPWINAVIN ----
SGT       QVGIVSWG Y--GCARPGYPG VYTEV STFASAIASAAR TL--
THR       QMGIVSWG E--GCDRDGKYG FYTHV FRLKKWIQKVID QFGE
```

Fig. 1. Sequence alignment of collagenase with human leukocyte elastase, thrombin and the six serine proteases used to build the original model. HLC (collagenase); HLE (human leukocyte elastase); ELA (elastase); CHT (chymotrypsin); TRP (trypsin); KAL (kallikrein); RMC (rat mast cell protease); SGT (*Streptomyces griseus* trypsin); THR (thrombin). The boxes correspond to the structurally conserved parts of the serine protease family (distance on Cα less than 1.6 Å), determined by the Greer method. The numbering is based on α-chymotryspin numbering.



(*a*) (*b*)

Fig. 2. (*a*) β-sheet 31–44 (yellow) shown in the superposition of MIR (magenta) and SIRAS (blue) electron densities, the red backbone corresponds to the model rebuilt in those maps. (*b*) β-sheet 31–44 shown in the final electron density. The strand on the left is 61–67 (see Fig. 6).

stant at 291 K. Crystallization was often hampered by twinning until NaCl was added. The crystals belong to tetragonal space group *I*422, $a = 111.7$, $c = 165.8$ Å. The crystal density was measured ($\rho_{crystal} = 1.2$) by centrifugation in a Ficoll gradient (Mikol & Giegé, 1992) and is consistent with two molecules in the asymmetric unit and 50% solvent.

2.2. *Preliminary structure determination*

Historically, an attempt to solve the structure by the multiple isomorphous replacement (MIR) method led us to screen many heavy-atom derivatives. Of the 25 heavy atoms tested (Arnoux, 1985), only $K_2Pt(CN)_4$ and $K_3UO_2F_5$ gave interpretable isomorphous Patterson difference maps, the latter also providing an anomalous map. The phasing power of heavy-atom derivatives was of poor quality because of the peculiar positions of both derivatives: on a twofold axis for the Pt atom and close to $x = 1/4$ for the U atoms. The quality of the MIR map was sufficient to determine the envelope, but not the orientation of the molecules. As HLC belongs to the trypsin family, an attempt was made to orient the molecules in the cell by the classical rotation translation methods (Crowther & Blow, 1967) using the coordinates for trypsin from the Protein Data Bank (Bernstein *et al.*, 1977) as a model. Presumably because of the high symmetry of the space group, this method failed.

At this step, a volume-recognition method was developed (Riche, 1985). The experimental MIR electron-density function is approximated by a three-dimensional Boolean function taking the value 'true' whenever above a threshold value. By subdividing in cubes with an edge of 2 Å, a sufficiently large parallelepipeded volume around the asymmetric unit, a logical function of integer $\rho(i,j,k)$ is defined. A 'model' function of integer $m(i,j,k)$ is generated in a similar way: a large cube including the whole macromolecule model (trypsin Cα backbone) is divided in small cubes with the same edge of 2 Å. A true value is set for $m(i,j,k)$ if a Cα atom is inside the cube with coordinates (i,j,k). The superimposition of $m(i,j,k)$ at a given position I,J,K on ρ allows the calculation of a figure of merit *i.e.* the recognition percentage which is equal to the sum of logical hits $m(i,j,k)$ and $\rho(I+i,J+j,K+k)$ on all cubes belonging to $m(i,j,k)$, divided by the maximum number of possible hits (number of Cα) and multiplied by 100. In a first step, a 'model' function composed of a filled sphere of radius 18 Å is translated along the three axes of the $\rho(i,j,k)$ function. The recognition translation function gave three peaks, one for each molecule and one in the middle. In the neighborhood of these peaks, a recognition rotation function $R(\omega,\chi,\varphi)$ is then calculated using the trypsin Cα model and an initial step angle of 10° on all three Eulerian angles ω, χ and φ. The orientation of both molecules of the asymmetric unit was determined unambiguously and shown to be related by a pseudo twofold axis [these solutions were later confirmed by *AMoRe* (Navaza, 1994) when this program became available, showing the efficiency of *AMoRe* to solve difficult cases]. At this early stage, rigid-body refinement was applied and gave a crystallographic R factor of 41% at 5 Å resolution. Starting from this solution, an attempt was made to refine the structure at 3 Å resolution with the diffractometer data, but this was not possible.

2.3. *Structure determination*

2.3.1. *Data collection and processing.* At this stage, two-dimensional area detectors became available and new data sets were recorded at high resolution. Data for the native protein were recorded at LURE using the prototype two-dimensional area detector Mark II (Kahn, Shepard, Bosshard & Fourme, 1996). Two data sets were recorded at 291 K on two different crystals. The data were processed with a local version of the *MADNES* program (Messerschmidt & Pflugrath, 1987; Bricogne, 1987) and integrated by an ellipsoidal masking procedure. In case of difficulty with autoindexing, the data were indexed in space group *P*1 with unit-cell dimensions $a = 111.7$, $b = 111.7$, $c = 114.5$ Å, $\alpha = 119.2$, $\beta = 119.2$, $\gamma = 90°$ and then transformed in *I*422. A global background array was initialized in the first 2° of data and updated as data collection proceeded. In addition the profile analysis of Kabsch (1988) was applied to the data as well as the usual Lorentz and polarization corrections. Details of the data collections are listed in Table 1. The two data sets were merged using programs from the *CCP*4 package (Collaborative Computational Project, Number 4, 1994) giving an R_{merge} of 3.9% at 1.7 Å resolution. The completeness of this data set is 74% on the totality of the data and 51% in the last resolution shell (1.8–1.7 Å).

The early heavy-atom data sets were recorded on a four-circle diffractometer but were only usable to 4 Å resolution. With the new two-dimensional area detectors available new data sets of the same heavy atoms [$K_3UO_2F_5$ and $K_2Pt(CN)_4$] were recorded to 2.4 Å resolution. In addition, there were several attempts to obtain a double derivative by soaking the crystal with both reagents. Many data sets were recorded, but only those that appeared to be usable will be presented here.

Two of the heavy-atom derivatives, one with platinum, and one with both platinum and uranium, were recorded using an Enraf–Nonius Fast system with an X-ray source running at 2.8 kW. In each case the crystal was pre-aligned to rotate about a crystallographic axis in order to record the Bijvoet pair on the same frame. Data were processed by using the *MADNES* program (Messerschmidt & Pflugrath, 1987) and integrated using a background plane-fitting algorithm followed by the profile analysis of Kabsch.

An additional platinum-derivative data set was recorded to 2.4 Å resolution on an MAR Research

Table 1. *Details of data collection for the native data sets*

M1 and M2 refer to the two native data sets recorded on the Mark II area detector at LURE.

	λ (Å)	D (mm)	Step (°/s)	2θ (°)	Resolution range (Å)	No. of observed reflections	No. of unique reflections	R_{sym} (%)	Completeness (%)
M1	1.3857	580	0.05/30	24	15.3–>1.7	101675	32496	4.6	54.2
M2	1.3790	580	0.05/30	24	15.6–>1.7	109495	42754	3.8	71.3

Table 2. *Details of data collection for the derivative data sets and details of MIR and SIRAS phasing quality*

The anomalous information of uranium was used only until 4 Å resolution. PtU, platinum and uranium derivative recorded on a Fast detector. Pt, platinum derivative recorded on an image-plate system.

(*a*) Data collection

	λ (Å)	D (mm)	Step (°/s)	2θ (°)	d_{max} (Å)	Unique reflections/ total reflections	R_{sym}* (%)	R_{anom}† (%)	Completeness (%)	R_{merge} (%)
Pt IPS	0.901	200	1.2/50	0	2.4	18834/ 53019	7.5	6.1	96	14.9
Pt Fast	1.5418	80	0.15/180	22	2.4	12754/ 35990	16.4	8.9	61	12.6
PtU Fast	1.5418	80	0.15/200	19	2.4	14789/ 49522	13.2	9.0	71	22.4

(*b*) MIR and SIRAS phasing quality

Method	Data	Site	Isomorphous occupation factor	Anomalous occupation factor	B_{iso} (Å^2)	R_{iso}§ (%)	R_{Cullis}¶ (%)	Phasing power** (%)	Figure of merit
MIR	PtU	Pt	1.03		18.27	11.65	65.7	2.81	0.67
		U	1.61		12.30				
		U	1.77		17.34				
	Pt	Pt	0.94		11.09	6.98	71.7	1.71	
SIRAS	PtU	Pt	0.80	1.57	35.24	14.37	58.9	2.11	0.69
		U	1.55	2.08	24.76				
		U	1.74	2.44	29.19				

* $R_{sym} = \sum_{hkl}\sum_{i=1}^{n}|F^2_{hkl,i} - \langle F^2_{hkl}\rangle|/\sum_{hkl} n\langle F^2_{hkl}\rangle$. † $R_{anom} = \sum_{hkl}\sum_{i=1}^{n'}|F^2_{hkl,i} - \langle F^2_{hkl,i}\rangle|/\sum_{hkl} n'\langle F^2_{hkl}\rangle$, n' represents the number of reciprocal lattice points equivalent in *I*422. ‡ $R_{merge} = \sum_{hkl}|F_{PH} - F_P|/\sum_{hkl} F_{PH}$. § $R_{iso} = \sum_{hkl}|F_{PH_{obs}} - F_{P_{obs}}|/\sum_{hkl} F_{P_{obs}}$. ¶ $R_{Cullis} = \sum_{hkl}|(F_{PH} \pm F_P) - F_H|/\sum_{hkl}|F_{PH} - F_P|$. Fom = $\langle\cos\Delta(\alpha_{best} - \alpha_i)\rangle$. ** Phasing power: $\langle|\Delta F|_{\pm h}\rangle/\langle|F''|\rangle = 2^{1/2}(N_A/N)^{1/2}\Delta f''_A/Z_{eff}$. N_A is the number of anomalous scatterers per molecule. N is the total number of non-H atoms in the molecule. Z_{eff} is the effective average atomic number ($\simeq$ 6.7 for proteins).

image-plate system located on the W32 beamline at LURE. The crystal was not pre-aligned as no automatic procedure for alignment was yet available in the processing program *MOSFLM* (Leslie, 1987). Furthermore, at the wavelength used (0.901 Å) the anomalous scattering for Platinum was rather poor. The data set was integrated also using *MOSFLM* by a profile-fitting method and then processed using the *CCP*4 package (Collaborative Computational Project, Number 4, 1994). The details of data collection are summarized in Table 2(*a*).

2.3.2. *Refinement of the model.* A model was built using the *Homology* program (BIOSYM). Using the Greer method of alignment (Greer, 1981, 1990), fragments from the structures of elastase, chymotrypsin, trypsin, kallikrein, rat mast cell protease and *Streptomyces griseus* trypsin were used by the program to build the model of the molecule. The alignment of the sequences of those six different proteins with that of HLC is presented in Fig. 1.

First of all, in order to verify the relative orientation of the two molecules of the asymmetric unit, this model was subjected to a rigid-body refinement using the program *X-PLOR* (Brünger, 1990). The refinement carried out between 10 and 4 Å resolution confirmed the predetermined orientation obtained in real space and gave an R factor of 42.67%. The model was then submitted to energy-minimization and slow-cooling cycles, using data greater than 3σ from 7 to 2 Å resolution leading to an R factor of 36% when refining with an overall B factor and 32% with individual ones. The non-crystallographic symmetry was imposed during all the procedure. The $2F_o - F_c$ map showed without ambiguity the conserved part of the structure of serine protease family but was excessively noisy and presented many discontinuities with no interpretable density in the difference map. We, therefore, concluded that some of the loops of the model were too far from their real position. *A posteriori* examination showed that the average phase error was 54° with an r.m.s. of 3 Å between the model and the final structure, which were too far to converge. The refinement

using our model as the only source of information had arrived at a dead end.

2.3.3. *MIR and SIRAS phasing methods.* At this stage the only possibility left to refine the structure was to use the MIR and/or SIRAS (single isomorphous replacement with anomalous scattering) method of phasing. A difference Patterson map was calculated for each derivative at 3 Å resolution. It confirmed the position of the heavy-atom sites previously determined (Arnoux, 1985). The positions were refined by the least-squares FHLE method (Dodson & Vijayan, 1971) and the phases refined by the program *PHARE* (Collaborative Computational Project, Number 4, 1994). The anomalous difference Patterson calculated using the U atoms was of poor quality but still indicated some usable information. Consequently the anomalous contribution was limited to 4 Å resolution in the estimation of SIRAS phases. Statistic on the phasing procedure are listed in Table 2(*b*).

2.3.4. *Rebuilding and refinement of the structure.* Those sets of phases led to the calculation of two electron-density maps at 2.7 Å resolution: a MIR map and a SIRAS map. Both were extremely noisy and discontinued, but when superimposed (Fig. 2*a*) with the last two maps ($2F_o - F_c$ and $F_o - F_c$) calculated after *X-PLOR*, it was possible to discern alternative tracings for the loops. Four reconstructions using *FRODO* were carried out on one molecule of the asymmetric unit, alternatively with *X-PLOR* minimization/simulated-annealing cycles between 8 and 2.5 Å using reflections greater than 3σ, followed by overall temperature-factor refinement. After each model-building session, transformations were applied from newly built molecule *A* to molecule *B* using the matrix calculated by *O* (Jones, Zou, Cowan & Kjeldgaard, 1991). New phases were combined with SIRAS phases leading to the calculation of a map weighted by the figure of merit. This map was superimposed on the two maps ($2F_o - F_c$ and $F_o - F_c$) calculated with the φ_{calc} phases derived from the atom coordinates. At the first stage, 41% of the residues were manually displaced, the main modification coming from the 'loop' 31–44, that appeared to be a β-sheet (Fig. 2*b*). The Cα atoms of residues 37*A*, 37*B* and 37*C*, that constitute the turn of the β-sheet, had to be displaced up to 15 Å. As this secondary structure enters deeply into the other molecule of the asymmetric unit, the refinement program would not have been able to modify this region, the previous location of those residues partly corresponding to the electron density of the β-sheet 31–44 of the other molecule of the asymmetric unit.

The refinement was carried on without increasing the resolution limit, until no further possible modification of the structure appeared in the maps. The resolution was then increased progressively by steps of 0.0125 Å^{-1} (proportional to the inverse of the shortest cell parameter), which corresponds to the maximum allowed value for phase extension. For later interventions, SIRAS phases were not used anymore and individual *B* factors were introduced in the refinement. Two additional global rebuildings of the molecule were necessary before reaching 2.3 Å resolution. Afterwards the two molecules of the asymmetric unit were rebuilt separately. At 2.3 Å resolution, water molecules were clearly visible in the electron density and introduced in the computation if thermal factors did not exceed a threshold value of 68 Å^2 (maximum thermal agitation factor of the main chain corresponding approximately to 2.7 times the mean temperature factor of the molecules) and if hydrogen-bond lengths were between 2.6 and 3.2 Å. At 1.88 Å resolution, no cut off in σ was used, because not enough reflections were retained using this criteria. The resolution limit of the refinement was kept at 1.8 Å. During the last step of the refinement, unambiguous water molecules were added even when having a high temperature factor. A summary of the refinement is shown in Fig. 3.

The quality of the map was sufficient to visualize an error in the chemical sequence: an inversion of Ile154 and Leu155. This assignment was later confirmed by gene sequencing. The sequence used in the final refinement is the one shown in Fig. 1. At the end of the refinement, 295 water molecules and 3556 atoms were introduced and led to a final *R* factor of 18.3% for 39 231 reflections. The quality of the final map was assessed by the calculation of a Fourier difference calculated with all atoms except one S atom which was used for calibration. The highest residual peak has a level of one fifth of that of the S atom. By removing 5.2% of reflections having $F_o/F_c > 2.5$ or $F_o/F_c < 0.4$ (Brzozowski, Derewenda, Dodson, Dodson & Turkenburg, 1992), the final *R* factor is 16.9%. The r.m.s. difference between the model built up with *Homology* and the final structure is 3.021 Å for the main chain. All the modifications took place in the non-conserved part of the serine protease family, apart from the last terminal helix residues that have been moved 20 Å away from the initial model. This difference

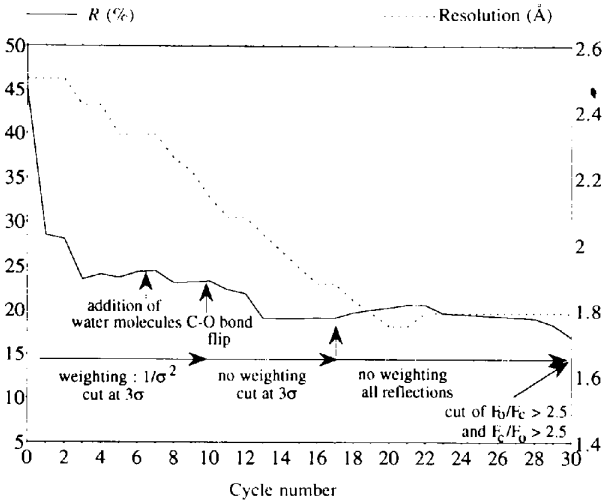


Fig. 3. Synopsis of refinement.

is explained by the presence of the last hydrophobic residue (Phe246) which needs to be sheltered from the solvent. Consequently, the helix is kinked at residue 242 and then turns toward the inner part of the molecule.

3. Results

3.1. *Quality of the structure*

In order to be able to easily compare HLC with other members of the serine protease family, the numbering of α-chymotrypsin has been adopted (16 to 246). In case of insertions, the residue number is followed by letters.

A Luzzati plot (Luzzati, 1952) was used to estimate the mean error in coordinates to be 0.20 Å. The fit of the structure to the density was evaluated by calculating a real-space *R* factor (Brandén & Jones, 1990) that is 0.81 for the main chain. The only parts of the structure which do not fit properly the density [real-space fit (r.s.f.) value lower than 0.7] are the residues Glu125*B* and Asp147 for both molecules, and the region Asp166–Pro173 for the *B* molecule only. The two first residues, for which even Cα has no clear density, are pointing toward the solvent channels around the fourfold axis that are alternatively 13 and 27 Å in diameter. It is to be noted that Asp147 corresponds to the position of thrombin autolysis (Bode, Turk & Karshikov, 1992) although no autolysis has been detected for HLC. The region Asp166–Pro173 corresponds to the extremity of a loop having a different environment in the two molecules. In molecule *A* this loop interacts with side chains from a symmetry-related molecule. In molecule *B*, it is in contact with the solvent and thus, not being constrained by hydrogen-bonding interactions, is shrunken around an enclosed water molecule.

Molecule *A* has a lower overall temperature factor (23 Å^2) than molecule *B* (26 Å^2) in spite of a slightly higher accessibility to solvent, 7504 Å^2 for *A* against 7167 Å^2 for molecule *B*. The *B* factor of HLC is generally proportional to the solvent accessibility for each residue, the maxima (68 and 52 Å^2) corresponding to the lowest values of r.s.f. for residues Asp147 and Glu125*B*, respectively. The only difference in temperature-factor behavior between the two molecules of the asymmetric unit is located on the loop Asp166–Pro173 for the reason described above.

The geometry of the structure is quite satisfying as shown by r.m.s. deviations (r.m.s. bond lengths = 0.015 Å, r.m.s. bond angles = 3.0°, r.m.s. dihedral angles = 27.7°, r.m.s. improper angles = 1.2°). The Ramachandran plots (Ramachandran, Ramakrishnan & Sasisekharan, 1963) of molecules *A* and *B* (Fig. 4) show no residues in the disallowed region. In each molecule, 19 of them are located in the allowed but not most favored region, six of which (Gln37*C*, Tyr74, Tyr99, Asn133, Asn207, Asn18) having a positive φ value. The first three residues are at the extremity of a type I turn, and Asn133 is at the third position of a turn of type II. Residue Asn207 is located immediately after the 3_{10} helix Ser202–Lys206 that acts like a turn for the β-sheet involving residues 198–201 and 208–215. Residue Asn18 is part of a small loop around a water molecule used to stabilize the N-terminal residue. The 13 other residues located in the limit of the allowed negative φ region, are mostly part of the substrate binding site, five of which having approximately the same φ/ψ values in all others serine proteases (Phe27, Trp41, Thr54, Asp102 and Ser214).

3.2. *Description of the molecular structure*

3.2.1. *Common features.* The two molecules of the asymmetric unit are related by a non-crystallographic

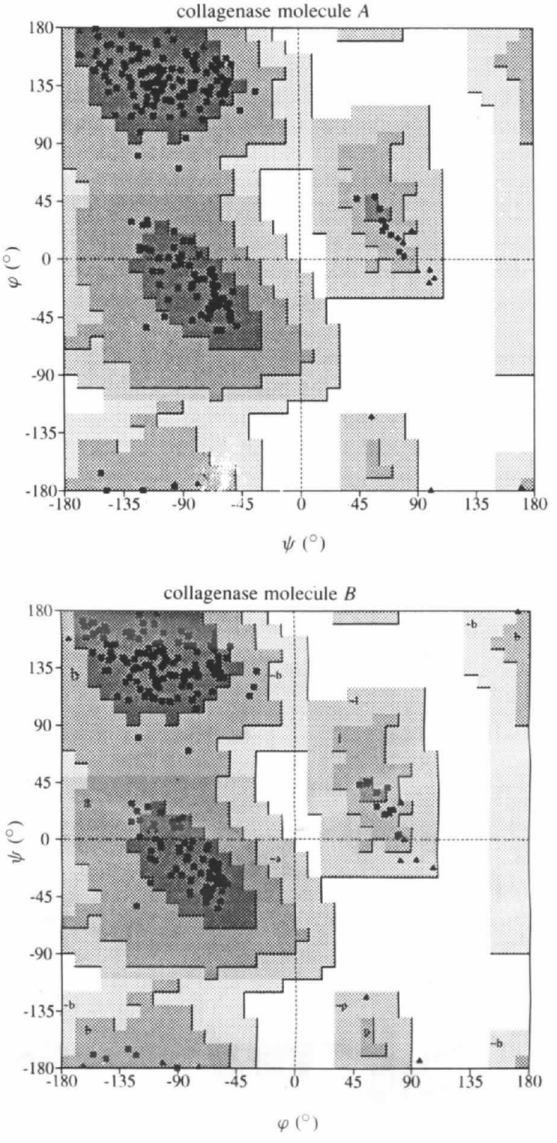


Fig. 4. Ramachandran φ,ψ plot for the two molecules of the asymmetric unit. Glycine residues are indicated by triangles.

twofold axis (Fig. 5). The two molecules can be considered as nearly identical as the r.m.s. difference between equivalent Cα atoms of the two main chains is 0.21 Å. Consequently, only the structure of molecule *A* will be described. The secondary structure of HLC (Fig. 6) is composed of two domains of the antiparallel β-barrel type containing eight and seven β-strands, respectively, related by a long loop of 26 residues (108–132) interrupted by a short one-turn 3_{10} helix (125*A*–129). HLC possesses three other one-turn helices also located at the surface of the protein, one of them (165–169) having an α-helical conformation, whereas the two others (202–206 and 56–59) are 3_{10} helices. The C-terminal helix starts as a 3_{10} helix at Val231, becomes an α-helix at Met235 and then is disrupted from Gly243 to the end in order to shelter Phe246 from the solvent. The number of disulfide bridges (which varies from two to four in the serine protease family) is limited to three in HLC. Two (42–58, 168–182) link a β-strand with an α-helix and the third one (191–220) links the extremity of two β-strands corresponding to the *P*1 and *P*2 binding sites (Polgár, 1987).

The two domains of HLC are connected by a β-sheet (20–22, 155–157) and by hydrogen bonds between residues 16–17 and two loops (143–147 and 189–194). Globally there are 42 hydrogen bonds between main chain and side chains, and 15 between side chains and side chains that stabilize the particular structure in a β-barrel of the two domains. In particular, Asp102 of the catalytic triad is hydrogen bonded to Ser229 and Ser214 which reinforces the link between the two domains. There are eight salt bridges in the HLC structure, two of them involving residues that are part of the β-sheet 31–44, and two others, both between Arg122 and Glu126 stabilizing the long loop separating the two domains.

3.2.2. *Detailed description of the asymmetric unit.* The main difference between the two molecules of the asymmetric unit concerns the side chain of Tyr20, the orientation of which differs considerably because of packing constraints. The OH of this tyrosine is linked to the main chain of a symmetry-related molecule for molecule *A*, but to a water molecule for molecule *B*, leading to a 180° rotation of the tyrosine plane around its Cα—Cβ bond. Otherwise, minor differences may be observed at position 167 for reasons already mentioned.

The intra-molecular hydrogen-bonding scheme of the two molecules is identical. In addition, there are two salt bridges connecting molecules of the cell, one between the two molecules of the asymmetric unit (Asp*A*60···Arg*B*38) involving β-sheet 31–44 and one with symmetry-related (–*y*, *x*, *z*) molecule (Glu*A*125*B*···His*A*111). The area of the protein surface which is removed from contact with the solvent is 2172 Å^2 with half of the contacts as a result of dimerization and half with symmetry-related pairs. So the two molecules are intimately embedded which causes low *B* factors for the atoms located at the surface of the upper barrel. The turn of the 31–44 β-sheet of one molecule of the asymmetric unit is hydrogen bonded with the residues of the catalytic pocket of the other molecule, thus blocking the entrance of substrates or inhibitors (Fig. 7).

3.2.3. *Solvent.* Water molecules are numbered following increasing values of temperature factor. Among the 295 water molecules fit in the electron density, 194 molecules are related by the pseudo-twofold symmetry, the others being equally distributed between molecule *A* and *B*. Among the common water molecules, ten have *B* factors lower than 25 Å^2 (mean temperature factor of the protein) with five molecules participating in the stabilization of the structure and five interacting with the active site. As the numbering is not the same for water molecules related by the non-crystallographic twofold axis, we will take as convention to indicate the two numbers separated by a slash, the first one corresponding to the water connected to molecule *A*.

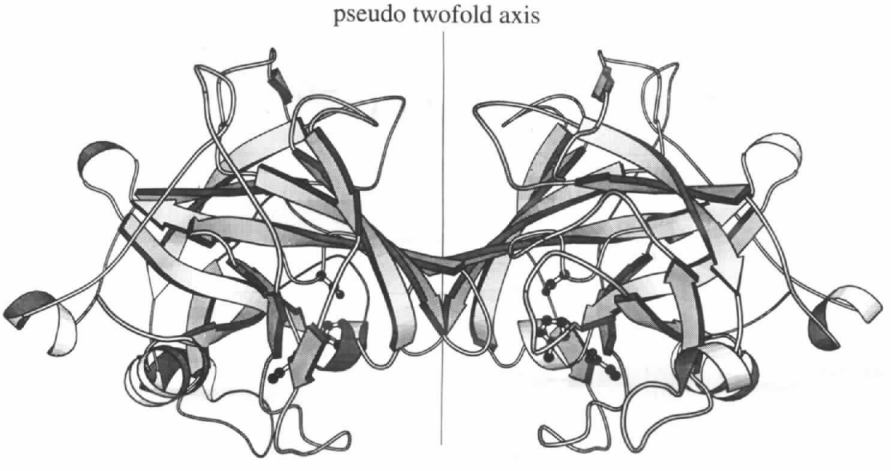


Fig. 5. *Molscript* (Kraulis, 1991) representation of the two molecules of the asymmetric unit of collagenase.

4. Discussion

4.1. *Tertiary structure*

The structure of HLC has been superimposed (see sequence alignment in Fig. 1) with human leukocyte elastase (HLE) (Bode *et al.*, 1986), pig elastase (ELA) (Meyer, Cole, Radhakrishnan & Epp, 1988), chymotrypsin (CHT) (Birktoft & Blow, 1972), trypsin (TRP) (Bode & Schwager, 1975), kallikrein (KAL) (Bode *et al.*, 1983), rat mast cell protease (RMCP) (Reynolds *et al.*, 1985), *Streptomyces griseus* trypsin (SGT) (Read & James, 1988) and thrombin (THR) (Bode *et al.*, 1989) using program *O*. The r.m.s. values are given in Table 3. CHT, KAL and RMCP crystallize with two molecules in the asymmetric unit but their relative orientation is different from HLC. The main structural difference between HLC and the other serine proteases is the presence of the 31–44 β-sheet, whereas a

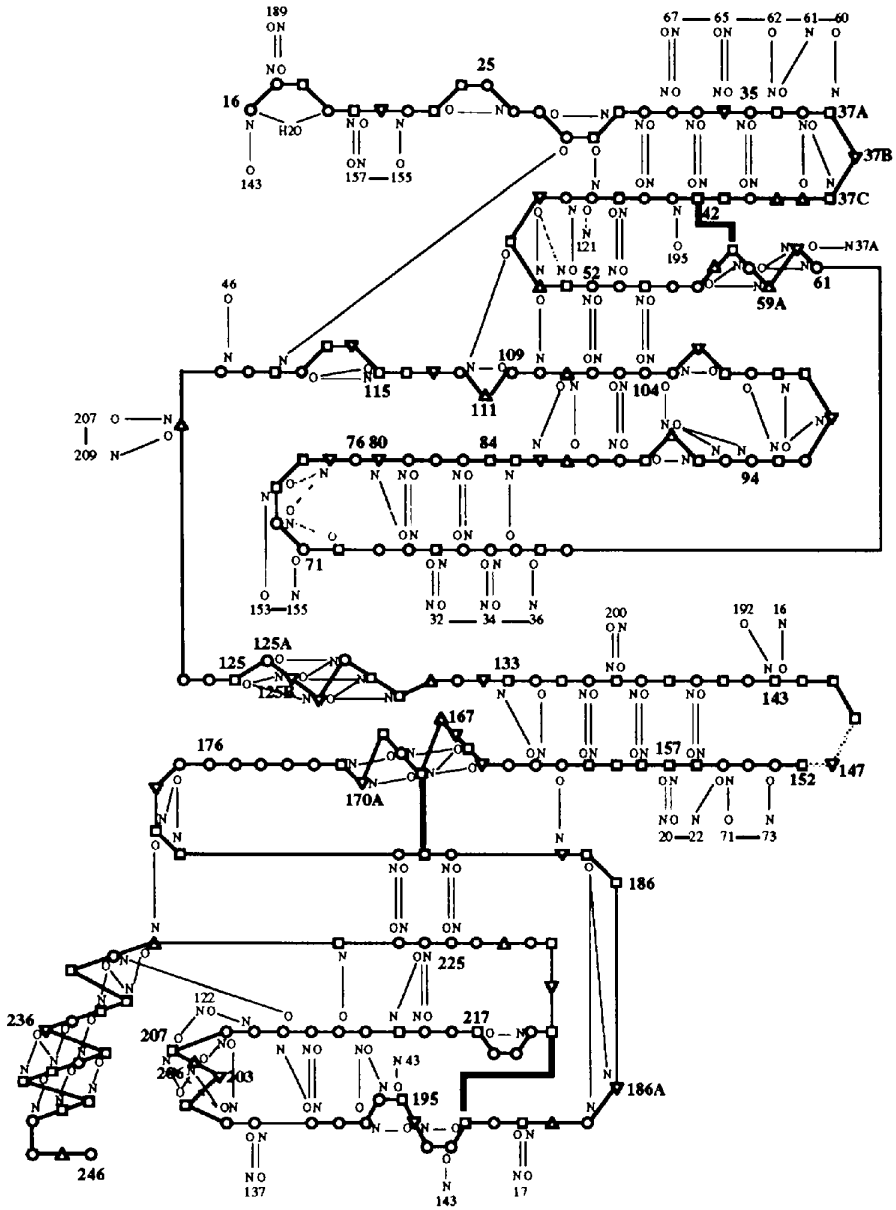


Fig. 6. Hydrogen-bond network formed between backbone atoms in collagenase from *H. lineatum*.

Table 3. *R.m.s. fit of molecule A of collagenase with seven serine proteases*

When two molecules are present in the asymmetric unit, the first one is used for the superposition.

	HLE	ELA	CHT	TRP	KAL	RMC	SGT	THR
Number of superposable Cα (distance < 3.8 Å)	194	203	199	200	196	205	187	195
R.m.s. fit on the 135 common superposable Cα (Å) (distance < 1.6 Å)	1.40	1.32	0.77	0.73	0.84	1.33	0.76	0.73
R.m.s. fit on the catalytic triad atoms (Å)	0.39	0.67	0.63	0.16	0.23	0.66	0.17	0.20

loop is found in all the others except RMCP. The β-sheet from RMCP is quite similar to the one observed in HLC but because of the different crystal packing, it does not interact with a symmetric counterpart. Pig elastase has a loop of the same length in this region (see sequence alignment and Fig. 8) which is almost perpendicular to the β-sheet of HLC. There are 34 hydrogen bonds, other than those imposed by the secondary structure, involving the β-sheet 31–44 (Table 4). Of those, 21 involve water molecules. It is stabilized by three side-chain/main-chain interactions and two salt bridges, one of which rigidifies the turn between the two strands. This β-sheet takes part in the secondary structure of HLC, forming a β-sheet with β-strand 42–46. It is also involved in the packing stability, making one hydrogen bond with a residue from the long loop separating the two domains of a symmetry-related molecule and six hydrogen bonds plus a salt bridge, with the other molecule of the asymmetric unit, mostly with residues involved in the binding site as will be discussed later.

4.2. *Active site*

For both molecules of the asymmetric unit, spatial relationship in the catalytic triad resembles the 'charge-relay system' first described (Birktoft & Blow, 1972) in α-chymotrypsin. The specificity pocket (Fig. 9), formed by Ser189, Val216 and Val226 is hydrophobic. Because of Val216, residues 189 and 226 are not accessible. A phenylalanine (Phe192), upstream of the consensus sequence GDSGGP (193–198), is located at the entrance of the binding pocket, restricting HLC recognition to small substrates or inhibitors. On the opposite side of the binding pocket is a tyrosine (Tyr99) which acts with Phe192 like a pair of tweezers, rendering the binding cavity almost flat. As already described, residues 37*B*–39 from molecule *B* are hydrogen bonded to residues involved in the binding pocket. Residues 37*B*–39 (DQRR) have five or more hydrogen bonds (Table 4), thus forming a complex network. As a consequence, the active site is completely blocked.

In Fig. 9 is also represented the CE—H···O=C interaction between His57 and Ser214 (3.13 and 2.98 Å for molecules *A* and *B*, respectively). This interaction has been reviewed by Derewenda, Derewenda & Kobos (1994) and can explain the particular φ/ψ angles of Ser214 (≃120/–55°) observed in all serine proteases.

Table 4. *Hydrogen bonds, others than secondary structure ones, between β-strand 31–44 and all molecules in the cell*

Number in parenthesis is the corresponding value with molecule *A*.

Atom	Residue	Number	Atom	Residue	Number	Distance (Å)
O	Ala	B31	NE2	Gln	B30	3.07 (3.05)
OD1	Asp	B34	OH2	Wat	4	2.63 (2.80)
OD2	Asp	B34	NE2	Gln	A143	3.09
OD2	Asp	B34	NH2	Arg	B39	2.70 (2.79)
OD2	Asp	B34	OH2	Wat	40	2.69 (2.61)
O	Asp	B34	OH2	Wat	8	2.89 (2.87)
OG1	Thr	B36	OH2	Wat	61	2.84 (2.76)
OG1	Thr	B36	OH2	Wat	8	2.78 (2.80)
O	Gln	B37A	OH2	Wat	173	2.84 (2.75)
OE1	Gln	B37A	NH2	Arg	B122*	3.27
N	Asp	B37B	OH2	Wat	103	3.16 (3.26)
OD1	Asp	B37B	N	Arg	B38	2.97 (3.27)
OD2	Asp	B37B	NH2	Arg	B38	2.69 (2.79)
OD2	Asp	B37B	OH2	Wat	48	3.11 (3.00)
O	Asp	B37B	OH	Tyr	A99	2.77 (3.08)
OE1	Gln	B37C	OH2	Wat	77	2.84 (3.02)
OE1	Gln	B37C	N	Val	A216	2.94 (2.86)
NE2	Gln	B37C	O	Val	A216	2.92 (2.97)
NE2	Gln	B37C	OH2	Wat	279	3.16 (3.06)
O	Gln	B37C	OH2	Wat	38	2.73 (2.78)
N	Arg	B38	OD1	Asp	B37B	2.97 (3.27)
NE	Arg	B38	O	His	A57	2.98 (3.22)
NH1	Arg	B38	O	Cys	A58	3.22 (2.95)
NH1	Arg	B38	OD2	Asp	A60	2.53 (2.59)
NH2	Arg	B38	OD2	Asp	B37B	2.69 (2.79)
N	Arg	B39	OH2	Wat	7	2.85 (2.96)
NE	Arg	B39	OH2	Wat	26	2.92 (3.20)
NH1	Arg	B39	OH2	Wat	26	2.70 (2.95)
NH1	Arg	B39	OH2	Wat	223	2.96
NH1	Arg	B39	OH2	Wat	262	2.62
NH2	Arg	B39	OH2	Wat	262	2.96
NH2	Arg	B39	OD2	Asp	B34	2.70 (2.79)
O	Arg	B39	OH2	Wat	52	2.82 (2.85)
O	Val	B40	OH2	Wat	4	3.13 (3.22)
NE1	Trp	B41	OH2	Wat	2	3.14 (3.11)
O	Trp	B41	OH2	Wat	9	2.67 (2.74)
O	Gly	B43	OG1	Thr	B54	2.90 (2.82)

* Symmetry-related $(-y, x, z)$ molecule.

There are four inner water molecules in the 'pocket' formed by the binding site and the residues 37*B*–39 from molecule *B* (Fig. 10). HLC has a water molecule (Wat38/23) at the location where ELA presents a sulfate ion. It lies in the so-called oxyanion site and interacts with OG of Ser*A*195 of the catalytic triad, with Arg*B*38, with Wat7/9 and with Trp*A*41. The serine is also hydrogen bonded to Wat230/99 and to Wat77/65, which is hydrogen bonded to His*A*57 and to Gln*B*37*C*. Wat7/9 is

also hydrogen bonded to Trp*A*41 as is Wat38/23 and to Arg*B*39.

4.3. *Inhibition modeling*

As no inhibitor of HLC other than DiFp is presently described, we tried to model other serine protease inhibitors in the HLC binding pocket. When compared with other serine proteases, only the elastase family has a valine at position 216 as in HLC, yielding specificity for smaller residues. Several pig elastase inhibitors have been tested with HLC, but none of them are active, indicating an incompatibility at the level of the *S*2 to *S*3 binding sites. For HLC, the consensus sequence GDSGGP (193–198) is flanked by two phenylalanines.

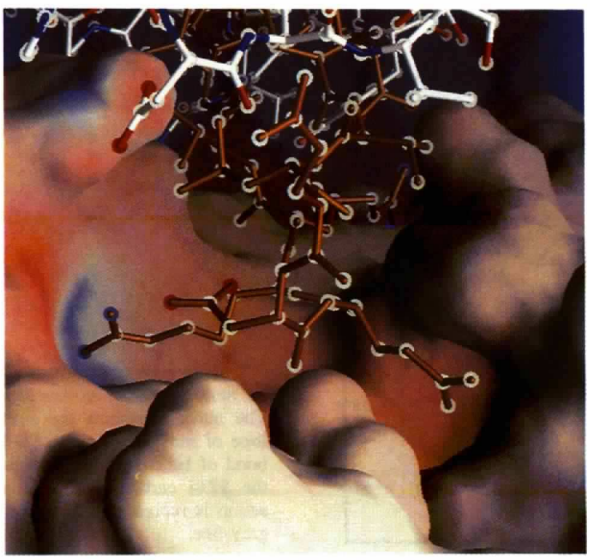


Fig. 7. Interface of the two molecules of the symmetric unit. The β-sheet 31–44 (in brown) of molecule *B* is represented by a ball and stick. The envelope of molecule *A* is represented by a white mask. The three penetrating residues of molecule *B* forming a tripod are D37B, Q37C and R38.

Phe199 is rather far away from the active site and found only in the thrombin family although not at the same spatial location as observed in the various thrombin X-ray structures, the aromatic ring of HLC Phe199 being stacked with Phe228 at a distance of 3.74 Å. Phe192, at the other extremity of the consensus sequence is close to the active site. It is only found in the sequence of myeloblastin (Bories, Raynal, Solomon, Darzynkiewicz & Cayre, 1989) and elastase (human, horse and murine). On the other side of the cavity, HLC has a tyrosine (Tyr99) whereas there is a valine in ELA and a leucine in HLE. This tyrosine is also observed in KAL but represents the only similarity between the HLC binding site and that of KAL. As the HLC cavity seems quite similar to that of HLE, the structure of HLE complex with meo-suc-Ala-Ala-Pro-Val-CH_2Cl has been superimposed to HLC to visualize the location of the inhibitor in its binding site. Apart from the clash with Tyr99, there is no other steric hindrance. As the phenolic side chain of the tyrosine is structurally free to swing out of its position, as had been observed in kallikrein upon binding BPTI (Chen & Bode, 1983), we may anticipate that this inhibitor will interact with HLC the same way as in HLE. The inhibitor seems to lay down on the surface of the molecule, its polypeptide chain being perpendicular to the β-sheet 31–44 axis and parallel to the phenyl group of PheA192. In order to characterize the inhibition of HLC by meo-suc-Ala-Ala-Pro-Val-CH_2Cl, activity tests will be performed.

The 31–44 β-sheet from molecule *B* mimics the way BPTI inhibits kallikrein (Gln37*C* is almost parallel to Lys15 of BPTI) in spite of the fact BPTI does not inhibit HLC. Because of the presence of a glycine and a serine at positions 216 and 226, respectively, in KAL, BPTI protrudes deeply inside the cavity. Residues 14–18 from BPTI have been superimposed with residues 37*B*–40 of the *B* molecule of HLC leading to an r.m.s. deviation of 0.47 Å for the main chain. The only noticeable difference is the flip of a carbonyl at position Cys14 in BPTI and Asp*B*37*B* in HLC. In the BPTI/KAL complex this carbonyl is in van der Waals contact with Met192. In

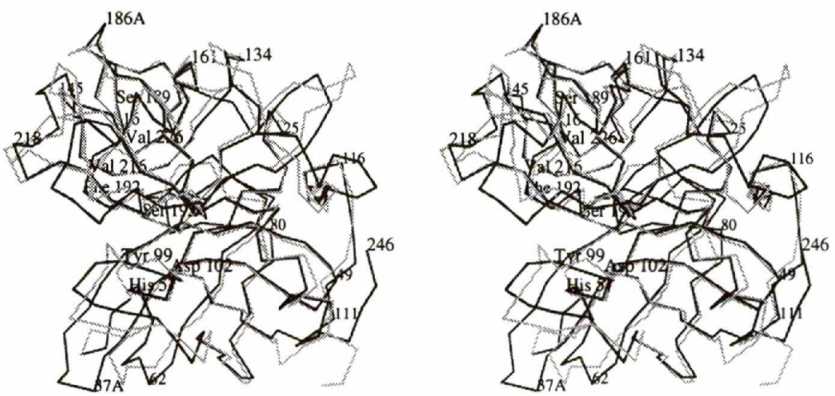


Fig. 8. Superposition in Cα trace of collagenase and pig elastase.

HLC, it is in contact with OH TyrA99, on the other side of the entrance to the cavity. This may prevent Tyr99 from rotating about its Cα—Cβ bound, keeping the entrance locked.

4.4. *Catalytic water molecule*

There has been much controversy about the water molecule involved in the tetrahedral intermediate (Perona, Craik & Fletterick, 1993). The water molecule Wat7/9 is a good candidate as it has a lower B value (20 Å^2) than the mean temperature factor of the protein and is found in all others serine proteases. The φ/ψ values of Trp41 are in the allowed but not most favored region of Ramachandran plot. This could be related to the hydrogen bond between the carbonyl of Trp41 and Wat7/9. This is observed in all serine proteases but does not imply a chemical role. Based on neutron experiments, another candidate for the hydrolytic water has been proposed by Singer, Smalås, Carty, Mangel & Sweet (1993), which would correspond to Wat77/65 (B = 34 Å^2) in HLC. Recently, HLC crystals were used to experiment a new technique for obtaining isomorphous derivatives (Schiltz, Fourme, Broutin & Prangé, 1995). Under moderate pressure xenon can bind to proteins with weak but specific interactions. The structure of the complex HLC/xenon has been determined, showing two Xe atoms, one in each binding site of the two molecules of the asymmetric unit. When superimposed with the HLC structure, no noticeable difference can be observed (r.m.s. = 0.12 Å on Cα, which is lower than the r.m.s. between the two molecules of the asymmetric unit in HLC). Only the hydration of the protein exhibits rearrangement, especially in the binding cavity. The Xe atom is at equal distance (3.5 Å) and is almost aligned with CH3 of Val216 and OH of Ser195. In the

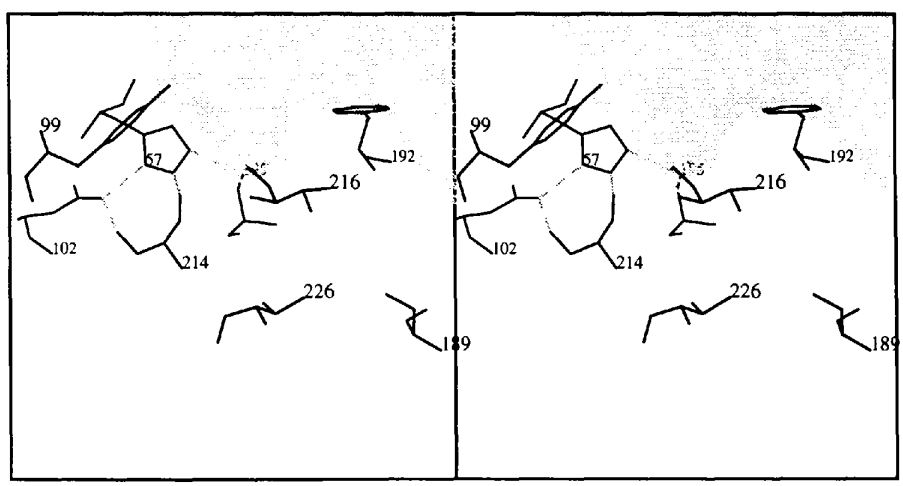


Fig. 9. Stereoview of the active site of molecule *A* of collagenase. The mask represents the envelope of molecule *B*. The C—H bond of H57 is directed toward the S214 carbonyl. This interaction is represented by a plain grey line.

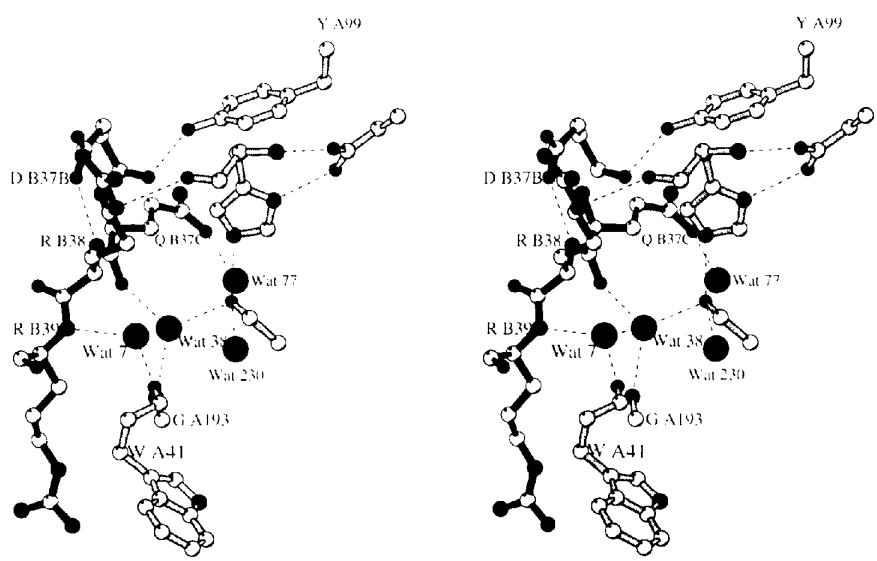


Fig. 10. Stereoview of the inner water molecules and the catalytic triad in the binding pocket of molecule *A* (in grey). Wat7, Wat38, Wat77 and Wat230 are hydrogen bonded to the binding pocket and to the β-sheet 31–44 of molecule *B* (in black).

HLC/xenon structure, Wat230/99 is no longer present and Wat77/65 is displaced 3 Å from its original position. It no longer forms a hydrogen bond with Ser*A*195 but does so with His*A*57 and Val*A*214 and is still linked to Gln*B*37*C*. The peculiar position of the Xe atom that mimics the tetrahedric intermediate geometry and the fact that Wat77/65 stays in the cavity favors the theory that this water molecule is the one involved in the enzymatic mechanism.

5. Concluding remarks

The crystal structure of HLC is a good example of the difficulties which are sometimes encountered when using molecular replacement. This is why we have detailed some of the technical difficulties which appeared during the process of rebuilding the molecule.

The structure shows a quite surprising auto-inhibition effect with one molecule active site being trapped by the other molecule of the asymmetric unit. The complexity and the number of hydrogen bonds implies a real recognition process.

Because the enzyme has been cloned and expressed recently, a structure/function study is now feasible.*

We thank C. Boulard for providing us with *H. lineatum* larvae. We are greatly indebted to R. Fourme, J. P. Benoît and R. Kahn for development of the DW32 (Fourme *et al.*, 1992) and D23 stations. We thank J. Navaza for running *AMoRe* with HLC data. We are grateful to J. P. Waller for critical reading of the manuscript. IB was supported by MRT. This project was supported by CNRS.

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1HYL, R1HYLSF). Free copies may be obtained through The Managing Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England (Reference: GR0443).

References

Arnoux, B. (1985). Thèse de doctorat, Université Paris XI, France.

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

Birktoft, J. J. & Blow, D. M. (1972). *J. Mol. Biol.* **68**, 187–240.

Bode, W., Chen, Z., Bartels, K., Kutzbach, C., Schmidt-Kastner, G. & Bartunik, H. (1983). *J. Mol. Biol.* **164**, 237–282.

Bode, W., Mayr, I., Baumann, U., Huber, R., Stone, S. R. & Hofsteenge, J. (1989). *EMBO J.* **8**, 3467–3475.

Bode, W., Reinemer, P., Hubert, R., Kleine, T., Schnierer, S. & Tschesche, H. (1994). *EMBO J.* **13**, 1263–1269.

Bode, W. & Schwager, P. (1975). *J. Mol. Biol.* **98**, 693–717.

Bode, W., Turk, D. & Karshikov, A. (1992). *Protein Sci.* **1**, 426–471.

Bode, W., Wei, A.-Z, Huber, R., Meyer, E., Travis J. & Neumann, S. (1986). *EMBO J.* **5**, 2453–2458.

Bories, D., Raynal, M. C., Solomon, D. H., Darzynkiewicz, Z. & Cayre, Y. E. (1989). *Cell*, **59**, 959–968.

Borkakoti, N., Winkler, F. K., Williams, D. H., D'Arcy, A., Broadhurst, M. J., Brown, P. A., Jonhson, W. H. & Muray, E. J. (1994). *Struct. Biol.* **1**, 106–110.

Boulard, C., (1970). *C. R. Acad. Sci. Ser. D*, **270**, 1349–1351.

Brandén, C.-I. & Jones, T. A. (1990). *Nature (London)*, **343**, 687–689.

Bricogne, G. (1987). *Computational aspects of protein crystal data analysis, Proceedings of the Daresbury Study Weekend*, pp. 120–145. Warrington: Daresbury Laboratory.

Brünger, A. T. (1990). *X-PLOR manual*, Version 2.1, Yale University, New Haven, Connecticut, USA.

Brzozowski, A. M., Derewenda, Z. S., Dodson, E. J., Dodson, G. G. & Turkenburg, J. P. (1992). *Acta. Cryst.* B**48**, 307–319.

Chen, Z. G. & Bode, W. (1983). *J. Mol. Biol.* **164**, 283–311.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Crowther, R. A. & Blow, D. M. (1967). *Acta. Cryst.* **23**, 544–548.

Derewenda, Z. S., Derewenda, U. & Kobos, P. M. (1994). *J. Mol. Biol.* **241**, 83–93.

Dodson, E. & Vijayan, M. (1971). *Acta Cryst.* B**27**, 2402–2411.

Ducruix, A., Arnoux, B., Pascard, C., Lecroisey, A. & Keil, B. (1981). *J. Mol. Biol.* **151**, 327–328.

Fourme, R., Dhez, P., Benoit, J.-P., Kahn, R., Dubuisson, J.-M., Besson, P. & Frouin, J. (1992). *Rev. Sci. Instrum.* **63**, 982–987.

Grams, F., Reinemer, P., Powers, J. C., Kleine, T., Pieper, M., Tschesche, H., Huber, R. & Bode, W. (1995). *Eur. J. Biochem.* **228**, 830–841.

Greer, J. (1981). *J. Mol. Biol.* **153**, 1027–1042.

Greer, J. (1990). *Proteins Struct. Funct. Genet.* **7**, 317–334.

Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.

Kabsch, W. (1988). *J. Appl. Cryst.* **21**, 67–71.

Kahn, R., Shepard, W., Bosshard, R. & Fourme, R. (1996). Unpublished work.

Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.

Lecroisey, A., Boulard, C. & Keil, B. (1979). *Eur. J. Biochem.* **101**, 385–393.

Lecroisey, A., Gilles, A. M., De Wolf, A. & Keil, B. (1987). *J. Biol. Chem.* **262**, 7546–7551.

Leslie, A. G. W. (1987). In *Computational aspects of protein crystal data analysis, Proceedings of the Daresbury Study Weekend*, 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.

Lovejoy, B., Cleasby, A., Hassell, A. M., Longley, K., Luther, M. A., Weigl, D., McGeehan, G., McElroy, A. B., Drewry, D., Lambert, M. H. & Jordan, S. R. (1994). *Science*, **263**, 375–377.

Luzzati, V. (1952). *Acta. Cryst.* **5**, 802–810.

Mandl, I. (1961). *Advance in Enzymologie*, Vol. 23, p. 163. New York: Interscience.

Messerschmidt, A. & Pflugrath, J. W. (1987). *J. Appl. Cryst.* **20**, 306–315.

Meyer, E., Cole, G., Radhakrishnan, R. & Epp, O. (1988). *Acta Cryst.* B**44**, 26–38.

Mikol, V. & Giegé, R. (1992). In *Crystallization of Nucleic Acids and Proteins: A Practical Approach*, edited by A. Ducruix & R. Giegé. Oxford: IRL Press.

Moiré, N., Bigot, Y., Periquet, G. & Boulard, C. (1994). *Mol. Biochem. Parasitol.* **66**, 233–240.

Navaza, J. (1994). *Acta. Cryst.* A**50**, 157–163.
Perona, J. J., Craik, C. S. & Fletterick, R. J. (1993). *Science*, **261**, 620–622.
Polgár, L. (1987). *Hydrolytic Enzymes*, Vol. 16, edited by A. Neuberger & K. Brocklehurst, p. 174. Amsterdam: Elsevier.
Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. (1963). *J. Mol. Biol.* **7**, 95–99.
Read, R. J. & James, M. N. G. (1988). *J. Mol. Biol.* **200**, 523–551.
Reynolds, R. A., Remington, S. J., Weaver, L. H., Fischer, R. G., Anderson, W. F., Ammon, H. L. & Matthews, B. W. (1985). *Acta Cryst.* B**41**, 139–147.
Riche, C. (1985). *A real space molecular replacement method.* Ninth European Crystallographic Meeting, Torino, p. 75.
Riès-Kautt, M. & Ducruix, A. (1992). *Crystallization of Nucleic Acids and Proteins*, edited by A. Ducruix & R. Giegé, pp. 195–218. Oxford: IRL Press.
Schiltz, M., Fourme, R., Broutin, I. & Prangé, T. (1995). *Structure*, **3**, 309–316.
Singer, P. T., Smalås, A., Carty, R. P., Mangel, W. F. & Sweet, R. M. (1993). *Science*, **259**, 669–673.
Spurlino, J. C., Smallwood, A., Carlton, D. D., Banks, T. M., Vavra, K. J., Johnson, J. S., Cook, E. R., Falvo, J., Wahl, R. C., Pulvino, T. A., Wendoloski, J. J., Smith, D. L. (1994). *Proteins Struct. Funct. Genet.* **19**, 98–109.
Stams, T., Spurlino, J. C., Smith, D. L., Wahl, R. C., Ho, T. F., Qoronfleh, M. W., Banks, T. M. & Rubin, B. (1994). *Struct. Biol.* **1**, 119–123.