Crystal structure of bovine procarboxypeptidase A-S6 subunit III, a highly structured truncated zymogen E

D.Pignol¹, C.Gaboriaud¹, T.Michon²,
B.Kerfelec³, C.Chapus³ and
J.C.Fontecilla-Camps^{1,4}

¹Laboratoire de Cristallographie et de Cristallogénèse des Protéines, Institut de Biologie Structurale, 38027 Grenoble Cedex 1, ²Laboratoire de Biochimie et Technologie des Protéines, INRA, Rue de la Géraudière, 44026 Nantes Cedex 07 and ³Laboratoire de Chimie Bactérienne du CNRS, Département de Bioénergétique et d'Ingénierie des Protéines, BP 71, 13402, Marseille Cedex 9, France

⁴Corresponding author

Communicated by C.-I.Brändén

Subunit III, a defective serine endopeptidase lacking the typical N-terminal hydrophobic dipeptide is secreted by the pancreas of ruminant species as part of the bovine ternary complex procarboxypeptidase A-S6. Two monoclinic crystal forms were obtained and subsequently used to solve its X-ray structure. The highest resolution model of subunit III was refined at 1.7 Å resolution to a crystallographic R-factor of 18.4%, with r.m.s. bond deviations from ideality of 0.012 Å. About 80% of the model presents the characteristic architecture of trypsin-like proteases. The remaining zones, however, have well-defined, unique conformations. The regions from residues 70 to 80 and from 140 to 155 present maximum distances of 16 and 18 Å relative to serine proteases and zymogens. Comparisons with the structures of porcine elastase 1 and chymotrypsinogen A indicate that the specific binding pocket of subunit III adopts a zymogen-like conformation and thus provide a basis for its inactivity. In general, the structural analysis of subunit III strongly suggests that it corresponds to a truncated version of a new class of highly structured elastase-like zymogen molecules. Based on the structures of subunit III and elastase 1, it is concluded that large concerted movements are necessary for the activation of zymogen E. *Key words:* activation process/protease E/serine endopeptidases/trypsin/X-ray structure

Introduction

Subunit III is a defective serine endopeptidase-like protein which has been found mainly in the pancreas of ruminants (Brown *et al.*, 1963; Puigserver and Desnuelle, 1975; Kerfelec *et al.*, 1985; Michon *et al.*, in press). In these animals, the protein is secreted in association with two pancreatic zymogens, the procarboxypeptidase A and a C-type chymotrypsinogen (Keil-Dlouha *et al.*, 1972). The resulting non-covalent ternary complex has been called proCPA-S6 since procarboxypeptidase A, which possesses two distinct binding sites for subunit III and the C-type chymotrypsinogen, is the central element of the association (Puigserver and Desnuelle, 1977; Granon *et al.*, 1990;

Michon *et al.*, 1991a). The presence of the ternary complex in the pancreatic juice of ruminants has been tentatively associated with some specific features of digestion in these animals (Kerfelec *et al.*, 1986; Michon *et al.*, 1991b); subunit III and the C-type chymotrypsinogen have been shown to protect procarboxypeptidase A, at least partially, against the acidic conditions in the duodenum of ruminants. A glycosylated form of subunit III, secreted both as a monomer (Guy-Crotte *et al.*, 1988) and as a binary complex in association with procarboxypeptidase A (Moulard *et al.*, 1989) has also been observed in human pancreatic juice. The proCPA-S6 ternary complex of ruminants can be reversibly dissociated under mild conditions into its constitutive subunits (Puigserver and Desnuelle, 1975; Kerfelec *et al.*, 1984) and the resulting free proteins have been extensively studied (for a review, see Chapus *et al.*, 1988).

Subunit III is made up of a single polypeptidic chain of 240 amino acids (Venot *et al.*, 1986). The protein contains five disulfide bridges (Venot *et al.*, 1986), four of them typical of pancreatic serine endopeptidases. The remaining one (Cys98 to Cys99b) should also be present in protease E, a separate member of the elastase family (Shen *et al.*, 1987) to which subunit III is closely related (86% amino acid sequence identity; Cambillau *et al.*, 1988). However, subunit III differs from protease E and other pancreatic serine endopeptidases in that it lacks the two hydrophobic N-terminal residues (Ile16/Val16 and Val17) characteristic of these molecules. Kinetic studies using *p*-nitrophenyl acetate as substrate (Kerfelec *et al.*, 1984) and ligand-binding experiments (Kerfelec *et al.*, 1986; Chapus *et al.*, 1990) suggest that the lack of specific activity of subunit III results mainly from a defective substrate-binding site that is only weakly functional (Wicker and Puigserver, 1981; Kerfelec *et al.*, 1984). This is not surprising since the two N-terminal hydrophobic residues are known to stabilize the enzyme active conformation through ion pairing of the α -amino group and Asp194 (Bode and Huber, 1986).

We have undertaken a crystallographic study of subunit III in order to understand the specific functional characteristics of this protein and to obtain new insights into the structural organization of pancreatic serine endopeptidases in general. Furthermore, the crystal structure of subunit III should provide the first three-dimensional model of a member of the protease E family. In this paper, we describe the three-dimensional structure of the bovine subunit III, as determined by X-ray crystallography using two crystal forms, and we define the structural basis for its inactivity. The general aspects of the activation process of serine endopeptidases are addressed by comparing the structure of subunit III with those of chymotrypsinogen A (Wang *et al.*, 1985) and elastase 1 (Meyer *et al.*, 1988). The structure of subunit III is also used to propose a description of the concerted movements occurring during the activation of zymogen E. A full account of the structural analysis and

crystallographic experimental details will be published elsewhere.

Results

Description of structure

The two different crystal forms of the bovine subunit III reported in this paper have been designated forms 1 and 2 based on their order of appearance. The standard crystallographic R-factor of the refined model of form 1 is 18.8% for data between 8.0 and 2.2 Å resolution and includes 120 water molecules. For form 2, which contains two molecules in the asymmetric unit, the R-factor is 18.4% for data between 8.0 and 1.7 Å resolution including 450 water molecules (Table I).

In none of the three independent models of the subunit III is there electron density for the first six to eight N-terminal residues. The X-ray model starts at Arg24 in form 1, and at Ser26 in form 2 (numbering according to the bovine chymotrypsinogen A scheme; Hartley and Shotton, 1971). All the remaining residues of the molecule are well defined (including those for which equivalent amino acids have been found to be disordered in zymogen structures). Figure 1 depicts a real space correlation plot (Jones *et al.*, 1991) corresponding to the two molecules in the asymmetric unit of the form 2 crystal. The corresponding correlations for the form 1 crystal are slightly lower (results not shown) which might be due to the intrinsic poorer quality of the crystals and/or to the lower resolution of the analysis. The three protein models from the two crystal forms are very similar. Comparison of the two molecules in the asymmetric unit of form 2 results in r.m.s. deviation of 0.41 Å when 230 Cα positions out of 232 well defined residues are considered. The r.m.s. deviation increases to 0.49 Å when N-terminal residues Ser26 and Trp27 are included in the calculation. Comparison of molecule A of form 2 with the refined model of form 1 results in an r.m.s. value of 0.54 Å for 221 residues. The value increases to 0.78 Å when all the Cα positions are considered. These values may be compared with the r.m.s. coordinate error of ~0.2 Å estimated from a Luzzati plot (Luzzati, 1952). The biggest differences between the atomic models of forms 1 and 2 are found in segments Asp116–Val118 and Ala203–Trp207 (with Cα–Cα distances ranging from 1.55 to 3.4 Å and from 1.99 to 4.4 Å, respectively). These segments, which

are located at the end of flexible loops, are stabilized by different packing contacts in the two crystal forms.

The overall polypeptide chain folding of subunit III indicates that it belongs to the trypsin family (Bode and Huber, 1986). The protein is composed of two hydrophobic interacting domains which participate in the formation of the catalytic site and the specificity pocket. In each domain, the chain forms long loops that fold as extended twisted β-pleated sheets. Only three segments are helical: a single turn (residues 55–59), a segment from residue 172 to 177 and a more extensive one comprising the C-terminal residues from 230 to 245.

Comparison with related structures

Only molecule A of the highest resolution form 2 will be used in the following discussions. However, since the three available models are very similar, the conclusions are generally valid.

We have chosen elastase 1 as the reference active enzyme structure because (i) it should be the closest known three-dimensional model (it has ~56% amino acid sequence identity with subunit III) and (ii) it has been solved to a similar resolution (Meyer *et al.*, 1988). The choice of chymotrypsinogen A as the zymogen reference structure has been dictated mainly by the degree of organization of its three-dimensional model (Wang *et al.*, 1985), superior to that of trypsinogen (Bode *et al.*, 1978). In addition, both subunit III and chymotrypsinogen A crystals have been obtained at around pH 4.5.

Superimposition of subunit III on elastase 1 and chymotrypsinogen A has been carried out using an iterative procedure that takes into consideration only pairs of Cα positions that are topologically equivalent. Under these circumstances, the r.m.s. difference between the reference active enzyme and subunit III is 1.04 Å for 195 out of the 232 well-defined Cα positions of the defective protein. A similar comparison of chymotrypsinogen A and subunit III results in an r.m.s. difference of 1.25 Å for 185 Cα positions, reflecting the lower degree of amino acid sequence homology (39%). From these results it is clear that, for ~80% of the structure, subunit III is very close to both reference molecules. The similarities include the loop containing the atypical disulfide bond (Cys98–Cys99b) and the catalytic triad. However, when all 232 well-defined Cα positions of subunit III are compared with their correspond-

Table I. Statistics on data collection and structure refinement for the two crystal forms (forms 1 and 2) of subunit III

	Form 1	Form 2
Space group	P2 ₁	P2 ₁
Molecules/asymmetric unit	1	2
Cell parameters	$a = 47.9 \text{ Å}, b = 61.1 \text{ Å}, c = 38.8 \text{ Å}, \beta = 95^\circ$	$a = 49.3 \text{ Å}, b = 82.7 \text{ Å}, c = 58.7 \text{ Å}, \beta = 97.9^\circ$
Resolution	2.2 Å	1.7 Å
Number of observations	21 180	124 312
Number of reflections	9207	44 851
Completeness (%)	81	93
R_{merge}	0.102	0.048
Final R-factor	18.8	18.4
Number of water molecules	120	450
r.m.s. bond deviations	0.014	0.012
r.m.s. angle deviations	1.61	1.37

ing counterparts in elastase 1 and chymotrypsinogen A, the r.m.s. deviations obtained are 4.5 and 9.3 Å, respectively. Thus, for a restricted part of its structure the defective protein presents large differences with both the active enzyme and the zymogen. These differences are mostly confined to four regions (Figure 2): (i) the Ser189–Ser195 zone of the activation domain; (ii) the loop comprising residues 68–78, known to represent the calcium-binding site in elastase (Dimicoli and Bieth, 1977) and trypsin (Bode and Schwager, 1975); (iii) the so-called autolysis loop, ranging from residues 142 to 155 and (iv) the region encompassing residues 116–119.

Differences between the activation domains

In subunit III, the Ser189–Ser195 segment is not very exposed to the solvent medium whereas in elastase 1 it protrudes into it (Figure 3A). One consequence of this difference is that the main chain atoms of Asn192 in subunit III occupy the site corresponding to the N-terminal Ile16 in elastase 1. The distance between Gly193 N and Ser195 N, the two nitrogen atoms of the oxyanion hole (which help stabilize the tetrahedral transition intermediate) is 4.3 Å in elastase 1 and 6.8 Å in subunit III. In addition, the whole region is rotated, resulting in a difference of 8.5 Å in the C α positions of the respective Gly193 residues. The overall r.m.s. difference for the Ser189–Ser195 segment is 4.5 Å.

Fig. 1. Correlation plot between the $(2F_o - F_c)$ electron density map and the models of the molecule A (thick lines) and B (thinner lines) of the second crystal form calculated with the program O. The correlation function is calculated including main chain and side chain atoms according to the following expression: $[1 - (\sum |e_{obs} - e_{calc}| / \sum |e_{obs} + e_{calc}|)]$.

Fig. 2. Distances between the C α positions after superimposition of subunit III and porcine elastase 1. The main differences are located in four regions: the calcium-binding loop (70–80), the 115–120 segment, the autolysis loop (140–155) and the specificity pocket (189–194).

The electron density map in this region, which is very well defined, is depicted in Figure 4. The other segments of the activation domain that line the specificity pocket (residues 213–220 and 225–228) have similar conformations in subunit III and elastase 1 although they are shifted relative to each other (up to 2.1 Å in the case of Gly219). These overall differences are similar to those found when comparing the three-dimensional structures of trypsinogen and chymotrypsinogen with those of trypsin and chymotrypsin (for more details, see Bode and Huber, 1986).

The conformations of the activation domains of subunit III and chymotrypsinogen A are similar, with an r.m.s. difference between the respective 185–195 loops of 1.3 Å (Figure 3B). However, the stabilization of Asp194 is different in the two structures. In subunit III, the carboxyl oxygens of this residue form hydrogen bonds with the main chain nitrogens of Arg143 and Trp141; in chymotrypsinogen A, Asp194 establishes a salt bridge with the imidazole ring of His40 (Bode and Huber, 1986) (Figures 5A and B). Despite these differences and as far as the activation domain is concerned, the subunit III model is clearly closer to a zymogen structure than to an active one. The stabilization of this part of the activation domain is obtained through interactions with the autolysis loop (residues 140–155) (Table II).

The calcium-binding site and autolysis loops. These are very well defined in subunit III (Figures 6 and 7) and present major differences with the corresponding regions in all known structures of active serine proteases and zymogens. The equivalent C α atoms of the calcium-binding site and autolysis loops of subunit III and elastase 1 have maximal distances of 16.2 and 18.5 Å, respectively (Figure 2 and 8). The differences in orientation are mostly modulated by residues Gly69, Gly78, Gly142 and Leu151 which represent the hinges of these flexible regions. These displacements also induce a 1.5 Å translation of the Glu36–His39 loop.

The calcium-binding loop of subunit III partially occupies a region corresponding to the main chain of residues Val23 and Pro24 of elastase 1. The stabilization of the conformation of this loop is mainly due to its interactions with the autolysis loop and the first three well-defined N-terminal residues (Table II). No calcium ion has been identified in the subunit III electron density maps. This is not too surprising since, in addition to the large differences in orientation, the corresponding calcium-binding site loops differ in eight out of 10 amino acids (although Glu70 and Glu80, to which calcium binds in elastase 1, are conserved).

The structural differences between the calcium-binding loops of the active enzyme and the defective protein are clearly correlated with differences on their autolysis loops. In subunit III, these two regions interact and are displaced in the same general direction relative to their positions in elastase 1. In fact, the autolysis loop displays a rather wide repertoire of conformations in the various serine endopeptidases: it is either completely or partially disordered in trypsinogen (Walter *et al.*, 1982) and chymotrypsinogen A (Wang *et al.*, 1985); it establishes few contacts with the rest of the molecule and is largely exposed to the solvent medium in elastase 1; it contributes to the stabilization of the activation domain in subunit III where it occupies the region corresponding to the calcium-binding loop of elastase 1.

Fig. 3. (A) Stereo view of the primary binding site specificity pocket in subunit III (thick lines) (segments Ser189–Ser195 and Val212–Cys220) optimally superimposed on the analogous segments of porcine elastase 1 (thin lines). (B) Superimposition of the same region with the homologous segment of chymotrypsinogen A (thin lines). The position of Gly193 in subunit III is closer to the one adopted in the zymogen structure than to the one adopted in the active enzyme.

Other differences in conformation

The distances between the C α positions of the Ser26–Lys30 segment of subunit III and corresponding amino acids of elastase 1 decrease progressively from 5.6 to 2.04 Å. In elastase 1, the N-terminal region (residues 16–25) approaches and partially penetrates the activation domain. This allows the N-terminal group of Ile16 to be stabilized by the side chain of Asp194 and the region to interact with the loop formed by residues 116–119 through a hydrogen bond between Gln119 NH and Pro28 CO. Because of the lack of salt-bridge stabilization in subunit III, the N-terminal region and the loop comprising residues 116–119 do not interact. Instead, this loop protrudes into the solvent medium and differs appreciably from the corresponding region of

elastase 1, with a maximal deviation of 6.0 Å for the C α of residue 118.

The overall conformational differences between subunit III, porcine elastase 1 and chymotrypsinogen A are summarized in Figure 8. In addition, a summary of the hydrogen bonds that stabilize these conformations is presented in Table II.

Discussion

Because of its observed non-specific residual activity, the three-dimensional structure of subunit III was supposed to have a somewhat active enzyme-type conformation and, consequently, to be closely related to porcine elastase 1

Fig. 4. Stereo representation of the $(2F_o - F_c)$ electron density map around the activation domain (residues 185–195) superimposed on the refined model. The electron density map was contoured at 1σ level.

Fig. 5. Stereo representation of the region around Asp194 in (A) subunit III and (B) chymotrypsinogen A.

Fig. 6. Stereo representation of the $(2F_o - F_c)$ electron density map around the calcium-binding loop (residues 70–80) superimposed to the refined model. The electron density map was contoured at 1σ level.

Fig. 7. Stereo representation of the $(2F_o - F_c)$ electron density map around the autolysis loop (residues 140–150) superimposed to the refined model. The electron density map was contoured at 1σ level.

(Kerfelec *et al.*, 1986). In fact, the high resolution X-ray structure of subunit III shows unexpected differences relative to all the other known trypsin-like serine protease and zymogen structures. It displays unique conformations at the calcium-binding site loop (residues 70–80), the autolysis loop (residues 140–155) and the 115–120 segment. These regions, separated in the primary sequence but spatially adjacent to the activation domain, play an important role in the activation process of serine proteases in general.

Structural bases of subunit III inactivity

During the activation of pancreatic zymogens, tryptic cleavage of the N-terminal activation peptide results in the emergence of an N-terminal hydrophobic dipeptide extremity which penetrates the activation cavity and forms an ion pair

through its α -amino group with the side chain of Asp194. The concerted rearrangement of the other neighbouring residues generates the correct structure of the binding site and of the oxyanion hole (Bode and Huber, 1986). Because the two hydrophobic N-terminal residues in subunit III are missing, the truncated N-terminal extremity cannot be stabilized in the same way. Instead, it seems to protrude into the solvent and be very mobile. As a consequence, the oxyanion hole and the specificity pocket are not properly organized and show a zymogen-type conformation, very similar to the one they adopt in chymotrypsinogen A (Figure 3B). Neither the region containing the atypical disulfide bridge nor the hydrogen bond network of the catalytic triad is different in the active enzyme and subunit III. It seems then that the inadequate conformation of the

Fig. 8. (A) Stereographic view of $C\alpha$ backbone in subunit III. Different colours were used to depict the regions presenting major differences with the known serine protease structures: red from Ser189 to Ser195; yellow from Gly140 to Leu155; green from Glu70 to Glu80; purple from Gly115 to Leu120 and white for the N-terminal region (Ser26–Ser32). Representation in the same orientation (B) for elastase1 and (C) for the chymotrypsinogen A. The same colours as in (A) are used for the different loops.

oxyanion hole and adjacent areas is solely responsible for the lack of specific activity.

A truncated zymogen E?

Comparison of the sequences of bovine subunit III and human protease E reveals a striking amino acid sequence

homology which includes the atypical disulfide bridge and leads to an identity score of 86%. However, since protease E displays specific catalytic activity and since it has the normal hydrophobic dipeptide sequence at the N-terminal end, its structure must be different from that of the truncated molecule. The conformation of subunit III, however, could

Table II. Hydrogen bond interactions between the calcium-binding loop (segment 70–80), the autolysis loop (segment 140–155) and the specificity binding pocket (segment 190–195) in bovine subunit III and porcine elastase 1

Subunit III	Porcine elastase 1
Tyr71 NH .. Pro28 CO	His71 Nε2 .. Asp21 Oε1
Tyr71 OH .. Trp141 Nε1	His71 CO .. Trp141Nε1
Arg73 NH .. Ser26 CO	Leu73 NH .. Gln153 CO
Ser74 CO .. Leu155 NH	
Leu76 NH .. Asp153 CO	
Glu80 Oε2 .. Leu151 NH	
Trp141 NH .. Asp194 Oδ1	
Gly142 NH .. Asn192 CO	Gly142 NH .. Asp194 Oδ1
Arg143 NH .. Asp194 Oδ2	Leu143 CO.. Val16 NH
Thr147 CO .. His40 NH	
	Ala152 NH .. Trp141 CO
Asp153 CO .. Gly69 NH	
	Asp194 Oδ1 .. Val16 NHter

The different stabilizations reflect the conformational modification.

Fig. 9. Activation process scheme in the protease E family. Subunit III (thick line) is used as a model for the truncated zymogen E. The dots represent the undefined N-terminal end. Elastase 1 (thin line) is used as the active enzyme model. The arrows describe the movements which occur in the course of activation after tryptic cleavage of the activation peptide and formation of the new N-terminal end.

be equivalent to the inactive precursor of protease E, the zymogen E (Kobayashi *et al.*, 1978), which differs from the active enzyme by the presence of an 11 amino acid activation peptide (Shirasu *et al.*, 1988). According to Pascual *et al.* (1990), it is the autolysis of zymogen E in the proCPA-S6 complex that gives rise to subunit III with the release of the first 13 residues, including the Val16–Val17 dipeptide. This suggests that the activation domains of zymogen E and subunit III are likely to adopt similar conformations since neither one can be occupied by the hydrophobic dipeptide. The most important implication of this observation is that subunit III may correspond to a truncated version of a new, highly structured class of zymogen molecules that includes zymogen E and, possibly, proelastase.

Activation process in the protease E family

If, as suggested above, protease E resembles elastase and subunit III is similar to zymogen E, the process of protease E activation should occur by a mechanism somewhat

different from the one described for trypsin and trypsinogen (Bode and Huber, 1986). According to these authors, the activation process takes place by the transformation of a flexible, easily deformable zymogen molecule into a more rigid, better organized active enzyme. The subunit III structure, however, clearly shows that in the case of the protease E family the inactive precursor molecule adopts a very well defined and stabilized structure. This implies in turn that the structural transition from zymogen E to protease E involves the passage from a highly structured inactive molecule to an active one of about the same level of organization. Although it is not possible at this point to determine the detailed sequence of the structural events in the course of activation, the reorganization of the molecule must occur via a series of concerted movements. Figure 9 depicts the various rearrangements that would be required for the activation to take place.

Subunit III in the procarboxypeptidase A-S6 complex

Subunit III is secreted by ruminant pancreas in association with procarboxypeptidase A and chymotrypsinogen C. In non-ruminant species procarboxypeptidase A is found either as a monomer (Guy-Crotte *et al.*, 1988; Moulard *et al.*, 1989) or, for example in the case of the porcine enzyme, associated with zymogen E (Martinez *et al.*, 1981). Tryptic activation of the porcine binary complex does not result in the release of protease E, suggesting that both zymogen and enzyme bind to the complex in the same manner. A consequence of this, and of the results presented above, is that it is unlikely that the regions implicated in the activation process participate in the complex association. We have recently obtained crystals of the proCPA-S6 heterotrimer. Structural analysis of this complex should shed more light on its function and lead to a detailed description of the interactions between their constituents. We have also initiated crystallographic analyses of zymogen E and protease E in order to confirm the hypothesis presented here.

Materials and methods

Purification

Subunit III was purified as previously reported (Kerfelec *et al.*, 1986) after the initial dissociation of the bovine ternary complex by dimethylmaleic anhydride (Kerfelec *et al.*, 1984).

Crystallization and data collection

Subunit III was crystallized in two different forms using the hanging drop vapour diffusion method, at 20°C. The first crystal form was obtained as previously described (Cambillau *et al.*, 1986), from a drop which initially contained 5 µl of protein solution (5 mg/ml) and 5 µl of well solution with 20 mM ammonium acetate, pH 4.2, and 20% saturated ammonium sulfate. Thin plate-like crystals grew up to 0.4 × 0.2 × 0.05 mm³ and diffracted to 2.2 Å resolution, using a synchrotron source (LURE, W32 station). The space group and cell dimensions were determined to be monoclinic, P2₁ with *a* = 47.9 Å, *b* = 61.1 Å, *c* = 38.8 Å, β = 95.03°, and with one molecule per asymmetric unit. Subsequently, the second crystal form was obtained using the same method, with a well solution of 100 mM ammonium acetate, pH 4.5, 25% (w/v) PEG 6000, and 8% (w/v) sodium chloride (Abergel *et al.*, 1991). The crystals, which grow as large prisms (0.7 × 0.7 × 0.5 mm³), contain two molecules per asymmetric unit and belong to the space group P2₁ with *a* = 49.3 Å, *b* = 82.7 Å, *c* = 58.7 Å, β = 97.9°. They diffract to 1.7 Å resolution.

X-ray intensity data were collected for both crystal forms, using the same synchrotron source on a MARsearch image plate scanner at 2°C. The data sets were processed with MOSFLM (Leslie, 1987) and scaled with the CCP4 crystallographic package (CCP4, 1979). For the first crystal form, a total of 21 180 reflections was collected from two different crystals, which resulted in 9207 unique reflections (corresponding to 81% of all data expected at

2.2 Å resolution) and an R_{merge} [defined as $\Sigma I(i) - \langle I \rangle / \Sigma I(i)$] of 0.109. For the second crystal form, 124 312 reflections collected for a final data set of 44 851 unique reflections (corresponding to 93% of all data expected at 1.7 Å resolution) and an R_{sym} of 0.048. Data statistics are summarized in Table I.

Determination of structure

A preliminary model for the first crystal form was obtained by molecular replacement (X-PLOR, Brünger *et al.*, 1987), using the porcine pancreatic elastase 1 structure as the initial model. Crystallographic refinement was performed using the simulated annealing and the energy minimization protocols of X-PLOR. The R-factor [defined as $\Sigma(|F_{\text{obs}}| - |F_{\text{calc}}|) / (\Sigma F_{\text{obs}})$], from 8 to 2.4 Å resolution including 90 water molecules was 0.24. In this model ~15% of the polypeptide backbone (the autolysis loop from Gly142 to Asp153 and part of the activation domain from 185 to 194) were not defined by clear continuous electron density in the map.

In order to know if those segments were really disordered, this model was then used as a basis for phase determination of the second crystal form by molecular replacement (AMoRe; Castellano *et al.*, 1992). The rotation and translation functions led to an unambiguous solution for the two molecules of the asymmetric unit, but with very bad contacts between them: the autolysis loop interpenetrated with its counterpart in the second molecule. The molecular replacement solution was confirmed by locating heavy-atom positions in a platinum salt heavy-atom derivative, previously obtained with a soaking time of 24 h in a solution containing 0.5 mM K_2PtCl_4 , 28% (w/v) PEG and 0.1 M ammonium acetate, pH 4.5. The protein was refined using the simulated annealing and the energy minimization protocols of X-PLOR with repetitive inspections of electron density maps using a graphic system. At the end of this process, the R-factor was 0.29 for data between 8.0 and 2.0 Å resolution. As for the model refined in the first crystal form, part of the structure was in weak density.

In order to improve the electron density definition, the automated refinement procedure (ARP) of Lamzin and Wilson (1993) was used. After 20 cycles of unrestrained refinement from 20 to 1.7 Å resolution, we obtained a much improved electron density map into which the Gly142–Asp153, Gly68–Gly78 and Gly185–Asp194 segments were manually rebuilt by using the graphics program O (Jones *et al.*, 1991). The first eight N-terminal residues seemed to be still disordered. This model was then refined using X-PLOR, with progressive addition of solvent molecules. The standard crystallographic R value for the 480 residues of the current model including 450 water molecules is 0.184 for 43 920 data from 8.0 to 1.7 Å resolution. The model has bond deviations of 0.012 Å and angle deviations of 1.37°. Examination of the Ramachandran plot showed three residues with (ϕ, ψ) angles outside the accepted region for non-glycine residues: Trp171 and Asp186 which appeared to be well-ordered, and Lys117, which is part of a loop in a region of weak electron density.

This model was then used to solve the structure of the first crystal form, which had been previously positioned in its unit cell using molecular replacement. After the X-PLOR refinement procedure, the R-factor for this form is 0.188 for data from 8.0 to 2.2 Å resolution and including 120 water molecules. The regions manually rebuilt into the electron density map generated by ARP were equivalent to, and as well defined as, the same parts of the three-dimensional model of subunit III in the second crystal form.

Structure comparisons

All the superimpositions were done with the program ALIGN (Cohen *et al.*, 1986) and displayed on a Evans & Sutherland ESV 20 graphic system using the program O (Jones *et al.*, 1991). The structures of porcine elastase 1 and chymotrypsinogen A were obtained from the Brookhaven Protein Data Bank (entries 3EST and 2CGA, respectively).

Acknowledgements

The assistance of Drs J.P. Benoit and R. Fourme of the W32 station of the LURE synchrotron radiation facility is gratefully acknowledged. The automatic procedure incorporated in Dr Victor Lamzin's program ARP was of great help in obtaining the refined model. We thank Dr Dominique Housset for encouragement and stimulating discussions. Dr S. Granon and E. Foglizzo are thanked for providing us with pure subunit III samples. D.P. was a recipient of a fellowship from the Ministère de la Recherche et de l'Espace.

References

Abergel, C., Moulard, M., Moreau, H., Loret, E., Cambillau, C. and Fontecilla-Camps, J.C. (1991) *J. Biol. Chem.*, **266**, 20131–20138.

- Bode, W. and Huber, R. (1986) In Desnuelle, P., Sjöström, H. and Norén, O. (eds), *Molecular and Cellular Basis of Digestion*. Elsevier, Amsterdam, pp. 213–233.
- Bode, W. and Schwager, P. (1975) *J. Mol. Biol.*, **98**, 693–717.
- Bode, W., Schwager, P. and Huber, R. (1978) *J. Mol. Biol.*, **118**, 99–112.
- Brünger, A.T., Kuriyan, J. and Karplus, M. (1987) *Science*, **235**, 458–460.
- Brown, J.R., Greenshields, R.N., Walsh, K.A., Yamasaki, M. and Neurath, H. (1963) *Biochemistry*, **2**, 867–876.
- Cambillau, C., Kerfelec, B., Foglizzo, E. and Chapus, C. (1986) *J. Mol. Biol.*, **189**, 709–710.
- Cambillau, C., Kerfelec, B., Sciaky, M. and Chapus, C. (1988) *FEBS Lett.*, **232**, 91–95.
- Castellano, E., Olivia, G. and Navaza, J. (1992) *J. Appl. Crystallogr.*, **25**, 281–284.
- CCP4 (1979) The SERC (UK) Collaborative Computing Project No. 4, a Suite of Programs for Protein Crystallography, distributed from Daresbury Laboratory, Warrington WA4 4AD, UK.
- Chapus, C., Puigserver, A. and Kerfelec, B. (1988) *Biochimie*, **70**, 1143–1151.
- Chapus, C., Kerfelec, B. and Dimicoli, J.L. (1990) *J. Biol. Chem.*, **265**, 3726–3730.
- Cohen, G.H., Satow, Y., Padlan, E.A. and Davies, D.R. (1986) *J. Mol. Biol.*, **190**, 593–604.
- Dimicoli, J.L. and Bieth, J. (1977) *Biochemistry*, **16**, 5532–5537.
- Granon, S., Kerfelec, B. and Chapus, C. (1990) *J. Biol. Chem.*, **265**, 10383–10388.
- Guy-Crotte, O., Barthe, C., Basso, D., Fournet, B. and Figarella, C. (1988) *Biochim. Biophys. Res. Commun.*, **156**, 318–332.
- Hartley, B.S. and Shotton, D.M.S. (1971) In Boyer, R.D. (ed.), *The Enzymes*. Academic Press, New York, Vol. 3, pp. 323–373.
- Jones, T.A., Zou, J.-Y., Cowan, S.W. and Kjeldgaard, M. (1991) *Acta Crystallogr.*, **A47**, 110–119.
- Keil-Dlouha, V., Puigserver, A., Marie, A. and Keil, B. (1972) *Biochim. Biophys. Acta*, **276**, 531–535.
- Kerfelec, B., Chapus, C. and Puigserver, A. (1984) *Biochem. Biophys. Res. Commun.*, **121**, 162–167.
- Kerfelec, B., Chapus, C. and Puigserver, A. (1985) *Eur. J. Biochem.*, **151**, 515–519.
- Kerfelec, B., Cambillau, C., Puigserver, A. and Chapus, C. (1986) *Eur. J. Biochem.*, **157**, 531–538.
- Kobayashi, Y., Kobayashi, R. and Hirs, C.H.W. (1978) *J. Biol. Chem.*, **253**, 2466–2470.
- Lamzin, V. and Wilson, K. (1993) *Acta Crystallogr.*, **D49**, 129–147.
- Leslie, A.G.W. (1987) *Acta Crystallogr.*, **A43**, 134–136.
- Luzzati, V. (1952) *Acta Crystallogr.*, **5**, 802–819.
- Martinez, C.M., Aviles, F.X., Sansegundo, B. and Cuchillo, C.M. (1981) *Biochem. J.*, **197**, 141–147.
- Meyer, E., Cole, G. and Radhakrishnan, R. (1988) *Acta Crystallogr.*, **B44**, 26–38.
- Michon, T., Granon, S., Sauve, P. and Chapus, C. (1991a) *Biochem. Biophys. Res. Commun.*, **181**, 449–455.
- Michon, M., Sari, J.C., Granon, S., Kerfelec, B. and Chapus, C. (1991b) *Eur. J. Biochem.*, **201**, 217–222.
- Michon, T., Kerfelec, B. and Chapus, C. *Protein Seq. Data Anal.*, in press.
- Moulard, M., Kerfelec, B., Mallet, B. and Chapus, C. (1989) *FEBS Lett.*, **50**, 166–170.
- Pascual, R., Vendrell, J., Aviles, F.X., Bonicel, J., Wicker, C. and Puigserver, A. (1990) *FEBS Lett.*, **277**, 37–41.
- Puigserver, A. and Desnuelle, P. (1975) *Proc. Natl Acad. Sci. USA*, **72**, 2242–2245.
- Puigserver, A. and Desnuelle, P. (1977) *Biochemistry*, **16**, 2497–2501.
- Shen, W., Fletcher, T.S. and Largman, C. (1987) *Biochemistry*, **26**, 3447–3452.
- Shirasu, Y. *et al.* (1988) *J. Biochem.*, **104**, 259–264.
- Venot, N., Sciaky, M., Puigserver, A., Desnuelle, P. and Laurent, G. (1986) *Eur. J. Biochem.*, **157**, 91–99.
- Walter, J., Streigemann, W., Singh, T.P., Bartunik, H., Bode, W. and Hubert, R. (1982) *Acta Crystallogr.*, **B38**, 1462–1472.
- Wicker, C. and Puigserver, A. (1981) *FEBS Lett.*, **128**, 13–16.
- Wang, D., Bode, W. and Huber, R. (1985) *J. Mol. Biol.*, **185**, 595–624.

Received on December 30, 1993