

A 2.3 Å resolution structure of chymosin complexed with a reduced bond inhibitor shows that the active site β -hairpin flap is rearranged when compared with the native crystal structure

Matthew R. Groves^{1,2}, Venugopal Dhanaraj^{1,4},
Mohammed Badasso¹, Philip Nugent¹, Jim E. Pitts¹,
Dennis J. Hoover³ and Tom L. Blundell^{1,4,5}

¹Laboratory of Molecular Biology, Department of Crystallography, Birkbeck College, Malet Street, London WC1E 7HX, ²Laboratory of Molecular Biophysics, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK, ³Department of Medicinal Chemistry, Central Research Division, Pfizer Inc., Groton, CT 06340, USA and ⁴Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge, CB2 1QW, UK

⁵To whom correspondence should be addressed

In the crystal structure of uncomplexed native chymosin, the β -hairpin at the active site, known as 'the flap', adopts a different conformation from that of other aspartic proteinases. This conformation would prevent the mode of binding of substrates/inhibitors generally found in other aspartic proteinase complexes. We now report the X-ray analysis of chymosin complexed with a reduced bond inhibitor CP-113972 {(2R,3S)-isopropyl 3-[(L-prolyl-*p*-iodo-L-phenylalanyl-S-methyl-cysteinyl)amino-4]-cyclohexyl-2-hydroxybutanoate} at 2.3 Å resolution in a novel crystal form of spacegroup R32. The structure has been refined by restrained least-squares methods to a final R-factor of 0.19 for a total of 11 988 independent reflections in the resolution range 10 to 2.3 Å. The extended β -strand conformation of the inhibitor allows hydrogen bonds within the active site, while its sidechains make both electrostatic and hydrophobic interactions with residues lining the specificity pockets $S_4 \rightarrow S_1$. The flap closes over the active site cleft in a way that closely resembles that of other previously determined aspartic proteinase inhibitor complexes. We conclude that the usual position and conformation of the flap found in other aspartic proteinases is available to native chymosin. The conformation observed in the native crystal form may result from intermolecular interactions between symmetry-related molecules in the crystal lattice.

Keywords: aspartic proteinase/inhibitor/X-ray structure/molecular replacement/chymosin

Introduction

Calf chymosin, an aspartic proteinase used for many centuries in the manufacture of cheese, is a bilobal neonatal gastric proteinase of molecular weight ~36 kDa consisting of 323 residues in its active form (Foltmann, 1970). The two lobes have a similar fold and are related by a pseudo twofold axis that runs perpendicular to the 25 Å substrate binding cleft (Tang, 1977). A catalytic aspartate in an Asp–Thr–Gly–Thr sequence is situated in a topologically equivalent position on each lobe of the enzyme.

Chymosin is secreted as an inactive precursor (prochymosin) and is activated at acidic pH by the proteolytic cleavage of the 42 residues of the N-terminal propeptide (Pedersen *et al.*,

1975). The main physiological function of chymosin is to cleave milk protein κ -casein, which acts as a stabilizer of the micelle (Jollés *et al.*, 1968), at an exposed Phe–Met bond. At the pH of milk (pH 6.6) chymosin cleaves the Phe105–Met106 peptide bond releasing the C-terminal peptide residues 106–169 known as glyco-macropeptide (GMP) which, in the presence of Ca^{2+} ions, results in the coagulation of whole milk micelles, leading to clotting (Raap *et al.*, 1983). Using circular dichroism and computer modelling techniques, it has been postulated that the κ -casein adopts an extended conformation in the region of the 105–106 bond and so could be easily accommodated within the substrate binding cleft of chymosin (Pedersen, 1977).

Chymosin is moderately specific to small substrates, as shown by its action on the B-chain of oxidized insulin (Pedersen and Foltmann, 1975); this implies that there exist interactions with the κ -casein substrate other than those within the active site cleft, which would account for the increased specificity. It is possible that a helical region of chymosin (residues 248–255, pepsin numbering) may form a recognition surface for the micelle system, although this has yet to be demonstrated.

The specificity of chymosin must be a consequence of the structure of the substrate binding pockets $S_4 \rightarrow S_3$ (Schechter and Berger, 1967; Lunney *et al.*, 1993), which have been identified by analogy with other high resolution mammalian, fungal and retroviral aspartic proteinase inhibitor complexes, such as renin (Dhanaraj *et al.*, 1992), pepsin (Chen *et al.*, 1992), penicillopepsin (James *et al.*, 1982), rhizopuspepsin (Parris *et al.*, 1992; Suguna *et al.*, 1992), endothiasepsin (Veerapandian *et al.*, 1990; Lunney *et al.*, 1993) and HIV-1 proteinase (Miller *et al.*, 1989; Priestle *et al.*, 1995). The S_1 and S_3 subsites form well defined but continuous pockets, accommodating hydrophobic residues; in chymosin the S_1 pocket is specific for large hydrophobic residues such as Phe. Residue Val111 in chymosin (pepsin numbering) was hypothesised to play a role in substrate recognition, lying as it does on the junction of the S_1 and S_3 pockets (Strop *et al.*, 1990). The cloning and expression of the enzyme, together with knowledge of its three-dimensional structure, now provide the opportunity of protein engineering new specificities for other peptide substrates that may have commercial applications in protein processing in the food or feed processing industries.

In designing such protein engineering experiments we need a crystal structure of the native and the inhibitor complexed enzyme. Surprisingly the native crystal structure (Gilliland *et al.*, 1990; Newman *et al.*, 1991) showed that the active site differed radically from that of other closely related mammalian aspartic proteinases, such as pepsin and renin. Indeed the extended β -hairpin known as 'the flap' has a similar extended conformation in all other aspartic proteinases for which there are crystal structures available. In the crystal structure of native chymosin the mainchain of the flap differs from this conformation and the conserved tyrosine (Tyr75) (pepsin numbering) is flipped by 180° around the C β –C γ bond so that

it occludes the space that would be expected to be occupied by the inhibitor or substrate at P₁; this movement and the consequences for self-inhibitory behaviour of chymosin were discussed by Andreeva *et al.* (1992). Furthermore all previous efforts to crystallize chymosin with inhibitors led to orthorhombic, inhibitor-free crystals identical to the native. These results posed two questions. Is the conformation observed in the native crystals characteristic of the native enzyme? Is the conformation of the enzyme in its complex with an inhibitor or substrate similar to that of other aspartic proteinases?

Recently, as a result of a systematic attempt to crystallize a series of chymosin inhibitor complexes, we obtained a novel rhombohedral crystal form of triangular prisms with the iodine-containing, reduced bond renin inhibitor (CP-113972). This iodine-containing inhibitor was the only synthetic inhibitor that produced inhibitor complex crystals. In this paper we describe the X-ray analysis at 2.3 Å resolution of these crystals. We describe the conformation of the inhibitor which lies in pockets S₄-S₁'. We compare the structure of the chymosin inhibitor complex with complexes of renin, pepsin and various fungal aspartic proteinases and show that the chymosin inhibitor complex is very similar. We compare the conformations of chymosin in the uncomplexed and complexed forms, and we discuss the nature and possible cause of the conformational change.

Materials and methods

Synthesis of CP-113972

Proton NMR spectra were obtained on a Varian Unity 400 spectrometer at 23°C. Liquid secondary ion mass spectra (LSIMS) were obtained on a VG70-250-S spectrometer using a liquid matrix consisting of 3:1 dithiothreitol/dithioerythritol. Microanalyses were performed by Schwarzkopf Microanalytical Laboratory (Woodside, NY). Chromatographic purification was performed using silica gel (30 µm) eluted with ethyl acetate-hexanes. Each substance was homogeneous by thin layer chromatography and ¹H NMR.

Preparation of (2R,3S)-isopropyl 3-[[N-(t-butoxycarbonylamino)-p-iodo-L-phenylalanyl-S-methyl-L-cysteineyl]amino]-4-cyclohexyl-2-hydroxybutanoate (Boc-p-I-Phe-SMeCys-norCStaOiPr) (1)

(2R,3S)-Isopropyl 3-[[S-methyl-L-cysteineyl]amino]-4-cyclohexyl-2-hydroxybutanoate (Hoover *et al.*, 1989; Hoover, D.J., Rosati, R.L. and Wester, R.T., manuscript in preparation) (2.54 g, 6.4 mM) and *N*-t-butoxycarbonyl-p-iodo-L-phenylalanine (1.0 equiv.) were coupled using triethylamine (1.1 equiv.), 1-(3-dimethylaminopropyl)-3-ethylcarbodiimide hydrochloride (DEC 1.0 equiv.) and 1-hydroxybenzotriazole hydrate (HBT, 1.5 equiv.) in dichloromethane at 0–25°C for 18 h giving, after extractive workup (ethyl acetate solution washed with aqueous HCl and NaOH) and chromatography (silica, ethyl acetate-hexanes), a colourless solid (3.15 g, 67%), homogeneous by TLC and ¹H NMR. LSIMS 734 (MH⁺). Anal. (C₃₁H₄₈N₃O₇SI) C, H, N.

Preparation of (2R,3S)-isopropyl 3-[[p-iodo-L-phenylalanyl-S-methyl-L-cysteineyl]amino]-4-cyclohexyl-2-hydroxybutanoate hydrochloride (2)

Boc derivative 1 (2.88 g, 3.92 mM) was dissolved in cold 4 M HCl-dioxanes (20 ml) and stirred at 23°C for 1.5 h. This solution was concentrated and dried, the residue ground up under ether, filtered and dried giving the deprotected product (2) as a light yellow solid (2.64 g, 100%). LSIMS 634 (MH⁺).

Preparation of (2R,3S)-isopropyl 3-[N-(t-butoxycarbonyl)-L-prolyl-p-iodo-L-phenylalanyl-S-methyl-L-cysteineyl]-4-cyclohexyl-2-hydroxybutanoate (3)

Hydrochloride (2) (1.01 g, 1.49 mM) was coupled with Boc-L-proline (1.1 equiv.), using triethylamine (1.1 equiv.), DEC (1.1 equiv.) and HBT (1.5 equiv.) according to the procedure used to prepare compound 1, and the crude product purified by chromatography giving a colourless foam (951 mg, 77%). LSIMS 831 (MH⁺).

Preparation of (2R,3S)-isopropyl 3-[[L-prolyl-p-iodo-L-phenylalanyl-S-methyl-L-cysteineyl]amino]-4-cyclohexyl-2-hydroxybutanoate (CP-113972)

Boc derivative 3 (0.95 g, 1.1 mM) was deprotected by the procedure described for the preparation of compound 2, except that the crude product was titrated with hot acetonitrile and dried, giving a colourless solid (685 mg, 81%): HPLC (Rainin Microsorb C-18, 250×4.6 mm, 1.0 ml/min, 60/40 acetonitrile/pH 2.1 0.1M KH₂PO₄ buffer) 5.1 min (95%); ¹H NMR (400 MHz, DMSO-*d*₆) δ 8.68 (d, 1H, J = 8.5 Hz), 8.39 (d, 1H, J = 8.3 Hz), 7.70 (d, 1H, J = 7.4 Hz), 7.57 (d, 2H, J = 8.3 Hz), 7.06 (d, 2H, J = 8.3 Hz), 5.33 (d, 1H, J = 6.2 Hz), 4.58 (m, 1H), 4.43 (m, 1H), 4.17 (m, 1H), 4.03 (m, 1H), 3.96 (dd, 1H, J = 5, 14 Hz), 2.68 (dd, 1H, J = 4, 10 Hz), 2.50 (dd, 1H, J = 9, 14 Hz), 2.23 (m, 2H), 2.04 (s, 3H), 1.84–1.70 (m, 4H), 1.62–1.50 (m, 4H), 1.35–1.28 (m, 2H), 1.23–1.00 (m, 4H), 1.14 (d, 3H, J = 6.2 Hz), 0.94–0.72 (m, 2H); LSIMS 731 (MH⁺). Anal. (C₃₁H₄₇N₄O₈SI.HCl) C, H, N.

Purification and crystallization

Commercial calf rennet, a gift of Dr A.Proctor (Pfizer, Groton, USA), was purified using a three-step purification process (Newman *et al.*, 1991):

- (i) concentration of crude rennet by ammonium sulphate precipitation;
- (ii) gel filtration using Sephadex G100 column;
- (iii) ion exchange on a monoQ FPLC column (Pharmacia).

Three peaks were eluted from the ion exchange column with relative milk clotting activities typical of A, B and C chymosins (Foltmann, 1970). The majority of the fraction used for crystallization was chymosin B, with little chymosin A; the remainder of the material (no more than 15%) consisted of chymosin C.

Chymosin inhibitor crystals suitable for X-ray analysis were grown using the hanging drop technique (McPherson, 1982) from a protein solution of approximately 10 mg/ml in 50 mM sodium phosphate containing 1.5 M NaCl at a pH of 6.0. The inhibitor was soaked in the protein solution in excess and filtered prior to crystallization experiments. Crystals began to appear after about one month and were fully grown after 2–3 months, having a triangular prism morphology with a maximum length of 0.3 mm.

Data collection and reduction

Data were collected at the SRS Daresbury (station 9.5) using a wavelength of 0.88 Å on a graphite monochromator to a resolution of 2.3 Å. 74° of data were collected on image plates. The data were indexed in spacegroup R32 with unit cell dimensions of a = b = 132.8 Å, c = 82.0 Å and processed using the MOSFLM (Leslie, 1993) suite of programs. The R_{merge} value for fully recorded observations was 0.095. The 62 777 observations of I > σ(I) were reduced to 12 125 independent reflections (11 348 acentric and 777 centric),

Table I. Summary of molecular replacement using native chymosin (pdb:4cms) (Newman *et al.*, 1991) as search model

Resolution limits of data used	10–3.5 Å
Spacegroup	R32
Sphere of integration in Patterson space	23 Å
Rotation function solution	$\alpha = 115.0^\circ$ $\beta = 153.4^\circ$ $\gamma = 336.8^\circ$
σ of rotation function	1.95
Peak height of rotation function solution	16.4 (8.5 σ)
Next highest peak in rotation function (discrimination of solution)	8.4 (4.3 σ)
Translation function solution (fractions of the unit cell)	$T_x = 0.47$, $T_y = 0.51$, $T_z = 0.58$
σ of translation function	1.41
Peak height of translation function solution	18.8 (13.3 σ)
Next highest peak in translation function (discrimination of solution)	11.0 (7.8 σ)
AMoRe correlation coefficient for the molecular replacement solution	0.615
AMoRe R-factor for the molecular replacement solution	0.385

giving a dataset which was 97.6% complete to 2.3 Å. Scaling of images was performed using the programs AGROVATA and ROTAVATA (CCP4, 1994).

Molecular replacement rotation and translation functions

Rotation and translation functions were calculated at a resolution of 3.5 Å, using the program AMoRe (Navaza, 1994) and the native co-ordinates (Newman *et al.*, 1991) (Protein Data Bank: 4cms), including the flap, as the search model. The rotation function was clearly interpretable with an 8.5 σ peak in the map at Eulerian angles $\alpha = 115.0^\circ$, $\beta = 153.4^\circ$, $\gamma = 336.8^\circ$. The molecular replacement solution had a correlation coefficient of 0.615 and an R-factor of 0.385. A more detailed analysis of the molecular replacement solution is given in Table I.

Refinement of the model

Initially four cycles of rigid body refinement were performed, with data in the resolution range 8.0–2.8 Å, using the least-squares refinement package RESTRAIN (Driessen *et al.*, 1989). This resulted in a rotation of 0.12° and a translation of 0.003 Å for the molecule treated as a rigid body, giving a correlation coefficient of 0.536 and an R-factor of 0.44 at 2.5 Å resolution. The model was then split into three rigid bodies (defined as residues –2–189, 190–302 and 303–327, pepsin numbering) and two further cycles of rigid body refinement were performed, after which the R-factor dropped to 0.40. A preliminary round of model building, using data to a resolution of 2.5 Å, was performed using Sim-weighted electron density maps with coefficients $m(2|F_o| - |F_c|)$ and $m(|F_o| - |F_c|)$ (Sim, 1959), which were displayed using FRODO (Jones, 1989). The electron density for conserved regions was generally encouraging, although density for surface loops 70→83, 156→163, 239→245 and 289→291 (pepsin numbering) was poor. Indeed negative density for Tyr75 was so strong that the co-ordinates of a pepsin inhibitor complex (Chen *et al.*, 1992) were used to aid model building for this region, being imported into the chymosin model and used as a framework on which to build the loop, before refinement against the reflection data. At this stage electron density for

the inhibitor was clearly identifiable, especially for the sulphur atom in P₂ and the iodine atom in P₃, which appeared as 4 σ and 5 σ peaks in the map. The inhibitor was built into the density in an extended conformation, although density for the P₄ Pro residue was initially ambiguous. Two further rounds of restrained refinement and model building were performed, after which the final R-factor for 11 988 reflections between 2.3 and 10 Å with $I \geq \sigma(I)$ is 0.19.

Results and discussion

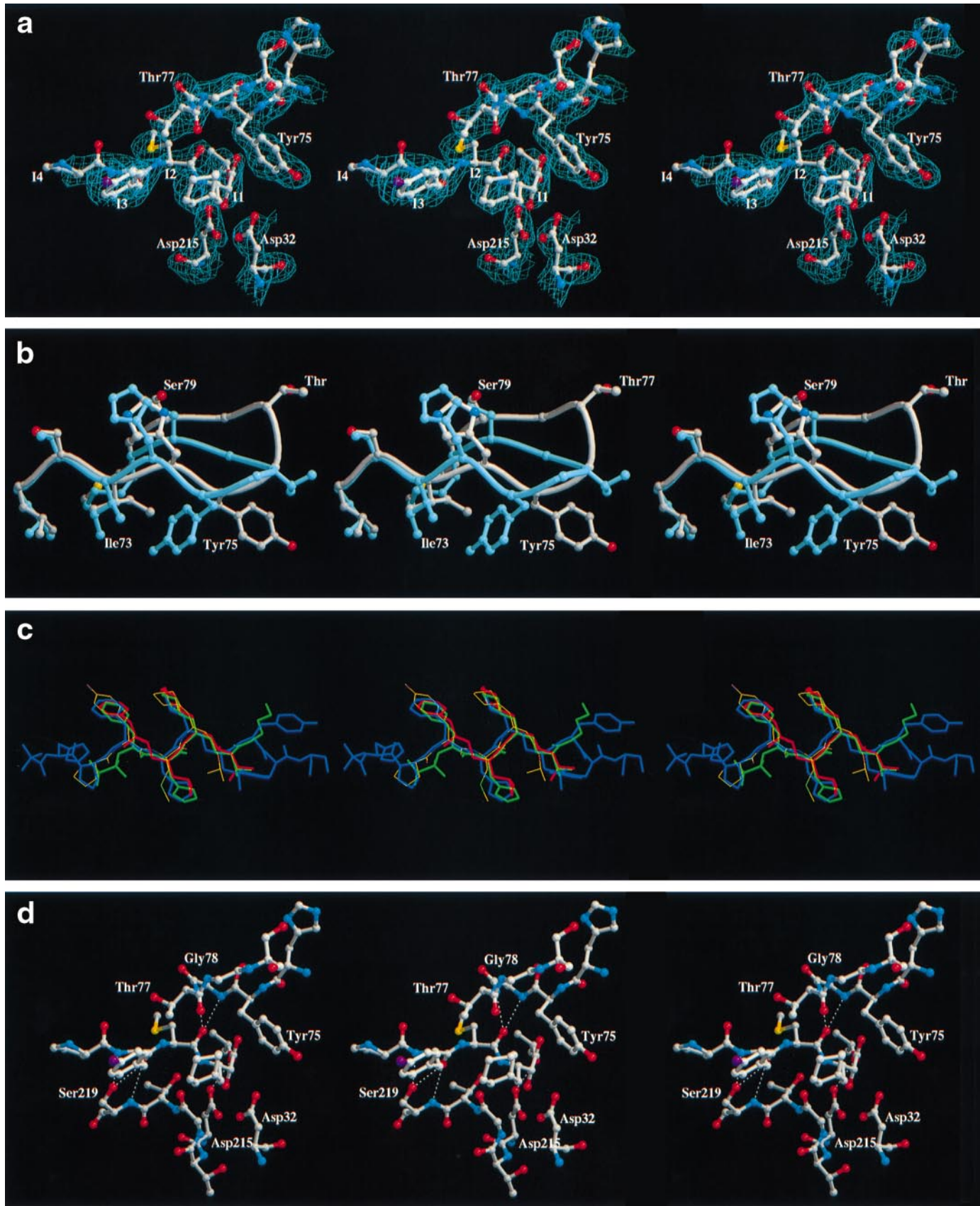
The chymosin inhibitor complex crystallized in a new space-group R32, cell dimensions $a = b = 132.8$ Å, $c = 82.0$ Å, with one molecule in the asymmetric unit and a solvent content of approximately 34%. This contrasts with the native enzyme, which crystallized in the spacegroup I222, cell dimensions $a = 79.7$ Å, $b = 113.8$ Å, $c = 72.8$ Å, with one molecule in the asymmetric unit and approximately 50% solvent content (Newman *et al.*, 1991). Molecular replacement and least-squares refinement led to a structure with good geometry and a final crystallographic R-factor of 0.19. The final electron density is illustrated in Figure 1a. The flap region 70–83 has reasonable electron density for all residues of the loop, while other surface loops such as 156–163, 239–245 and 289–291 still show high temperature factors in the complex, with breaks in density for the loop 289–291. In hexagonal porcine pepsin (Cooper *et al.*, 1990) and native chymosin (Newman *et al.*, 1991) these residues are also disordered, although they have crystallized in a different spacegroup. This suggests that the disorder is due to conformational variability rather than a consequence of the crystal packing. The disorder in the region 239–245 may be a consequence of sequence differences in this region resulting from a mixture of chymosin A, B and C (Foltmann, 1970).

The final model was aligned with the uncomplexed structure and rigid body shifts identified using the program XS5 (Säli, A., unpublished program); as shown in Table II. The tertiary structure is characteristic of previously solved structures of aspartic proteinases; r.m.s. deviation comparisons with other inhibitor complexes are given in Table III.

Aspartic proteinases show a rigid body movement of residues 190–302 (pepsin numbering) when complexed with inhibitors and even in differing crystallographic environments (Säli *et al.*, 1989, 1992). When compared with the pepsin inhibitor complex (pdb:4pep; Chen *et al.*, 1992) where the domain movement is very large it is noticeable that there is a smaller domain movement upon complexation for the chymosin complex, which accounts for the poor degree of similarity between chymosin and pepsin inhibitor complexes. Although both complexes show the inhibitor bound in an extended conformation, the inhibitor in the chymosin complex occupies a greater volume of space within the active site cleft than the inhibitor in the pepsin complex; hence there is no requirement of the protein to undergo large domain movements to minimize the volume of unoccupied space. Most of the differences in the positions of residues in the active site cleft result from local conformational changes due to the presence of the inhibitor, with the largest being that of the β -hairpin over the active site (Figure 1b).

Inhibitor–protein contacts that define the specificity pockets $S_1 \rightarrow S_4$

This is the first chymosin inhibitor complex to be reported, although many attempts have been made to co-crystallize



various inhibitors with chymosin. The inhibitor lies in an extended conformation within the binding cleft making numerous electrostatic and hydrophobic interactions as in other aspartic proteinase complexes (Figure 1c). The catalytic aspartates are found in the expected coplanar conformation with the OH of the inhibitor lying in the plane of Asp32 and Asp215. A network of hydrogen bonds is formed around these essential residues, known as the 'fireman's grip' (Pearl and Blundell, 1984), which holds the aspartates in this rigid conformation. In inhibitor complexes of other aspartic proteinases (Dhanaraj *et al.*, 1992) the Asp215 is usually found closer to the OH of the inhibitor. Both the carbonyl oxygen of Gly34 and the O δ 1 of Asp215 are within hydrogen bonding distance of the oxygen of the O₁C₃H₇ moiety of the inhibitor (3.4 Å, 2.9 Å respectively). The extra hydrogen bond available to Asp215 in the chymosin inhibitor complex may account for this difference (Figure 1d).

The flap region 71–81 is rearranged when compared with that of the native, accommodating the large cyclohexyl ring of P₁. Gly76 makes a hydrogen bond between its nitrogen to

the OG oxygen in the inhibitor (3.5 Å) which, along with the hydrogen bond from Trp39 to Tyr75, appears to stabilize the flap in this closed conformation.

The S₁' pocket lies between the tip of the flap (residues 70–76) and the active site (residues 30–34 and 215–218), with contributions from residues 112–120; it is predominantly non-polar and accommodates the isopropyl ester moiety (O₁C₃H₇). The cyclohexyl ring at P₁ occupies the S₁ pocket and is surrounded by the aromatic rings of Trp39, Tyr75, Phe112, Phe117 and the aliphatic sidechains of Ile120 and Leu30. The cyclohexyl ring also contacts the para-iodo-phenylalanine (pIPhe) at P₃, and thus is totally enclosed within a hydrophobic pocket, suggesting that this pocket is specific for a large, hydrophobic residue at P₁.

The specificity pocket S₂ is again comprised mainly by non-polar residues, although Thr77 and Thr218 make stabilizing contributions to the S γ and N of P₂. This pocket is much more open around the S-methyl cysteine (SMC) of P₂ and could easily accommodate a larger residue, such as His, Phe or Met. P₂ contributes to some of the hydrophobic packing around the O₁C₃H₇ moiety at the C-terminus of the peptide inhibitor; which partially occupies S₁'.

S₃ is a large pocket, which accommodates the pIPhe P₃ residue. The main chain of P₃ makes contact with Ser219 (O–219 N:2.8 Å, N–219 O γ :3.2 Å). The pocket is not very tight around P₃ allowing for a degree of movement of this residue; this may account for the incomplete pIPhe ring density in the final maps.

The S₄ specificity pocket lies towards the end of the binding cleft, so that P₄ Pro makes few interactions with protein atoms, although there are several well-ordered water molecules in this region. The P₄ Pro appears to have a large degree of freedom for rotation around the peptide bond with P₃, leading

Table II. Rigid body shifts of the inhibitor complex with respect to the native enzyme

	Rotation (°)	Translation (Å)
N-terminal domain (–2 to 189) (excluding the active site flap)	0.95	0.06
Active site flap (71–81)	27.13	0.11
C-terminal domain (190–322) (excluding helix 248–255)	2.75	0.02
N-terminal lobe (303–327)	0.96	0.25
Helix 248–255	1.67	0.59

Table III. Structural alignment of chymosin–inhibitor complex with other aspartic proteinases

	R.m.s. (Å)	Average difference (Å)	No. of structural equivalences	Poorly aligned regions (pepsin numbering)
Recombinant human renin complex ¹	1.06	0.96	242	144–147,156–162,233–254,278–282
Mouse submaxillary renin complex ²	1.08	0.98	264	142–147,187–204,233–244,280,289–299
Porcine pepsin complex ³	0.90	0.79	292	56–159,236–244,275–280,289–297
Endothia <i>parasitica</i> complex ⁴	0.90	0.79	198	7–10,46–52,156–161,183–188,194–211,222–244, 249–282,286–297,316–319
Rhizopus <i>chinensis</i> complex ⁵	1.44	1.15	263	48–52,108–112,156–161,185–181,200–209,238–244, 249–254,275–291
Bovine chymosin ⁶	0.69	0.61	303	74–77,156–161,289–297
Human recombinant renin ⁷	0.97	0.86	272	144–147,156–162,199–204, 240–244,276–282,289–297
Porcine Pepsin ⁸	0.91	0.81	302	156–160,290–297
Endothia <i>parasitica</i> ⁹	0.98	0.82	195	46–52,142–145,156–161,184–211,223–259, 263–284,316–319
Rhizopus <i>chinensis</i> ¹⁰	1.45	1.16	261	109–112,156–161,185–188,200–203,238–244, 249–254,274–281

¹pdb:1rne (Raheul *et al.*, 1991); ²pdb:1smr (Dealwis *et al.*, 1994); ³pdb:1psa (Chen *et al.*, 1992); ⁴pdb:2er7 (Veerapandian *et al.*, 1990); ⁵pdb:4apr (Suguna *et al.*, 1992); ⁶pdb:4cms (Newman *et al.*, 1991); ⁷pdb:2ren (Sielecki *et al.*, 1992); ⁸pdb:5pep (Cooper *et al.*, 1990); ⁹pdb:4ape (Pearl and Blundell, 1984); ¹⁰pdb:2apr (Suguna *et al.*, 1987).

Fig 1. (a) A figure showing the electron density for the bound inhibitor CP-113972 and the flap (residues 71–80) in the closed conformation (contoured at 1.2 σ). (b) A figure comparing the conformation of the flap (residues 71–80) in chymosin/CP-113972 complex and native chymosin (Newman *et al.*, 1991) (light blue). Tyr75 rotates by approximately 180° around the C α –C β bond. (c) A figure showing the superposition of CP-113972 with previously determined inhibitors of renin (Tong *et al.*, 1995; Dhanaraj *et al.*, 1992). CH-66 is shown in blue, CGP 38'560 is shown in green and BILA 980 is shown in red. (d) A figure indicating the hydrogen bonds made between CP-113972 and residues of the flap and the active site of chymosin.

to relatively poor density. Several residues make contributions to two pockets simultaneously; for example, Val111 makes contributions to specificity pockets S_1 and S_3 and has already been the subject of site-directed mutagenesis studies, showing it to have an effect on the k_{cat} of this enzyme (Strop *et al.*, 1990). Ile300 lies between the sulphur of the *S*-methyl cysteine (SMC) P_2 and the isopropyl ester moiety of the P_1 norstatine (Nor), which partially fills the S_1' pocket, and may play a part in keeping these two regions apart, producing the extended conformation in the P_1' - P_2 region.

Conformational changes within the binding cleft

The co-ordinates of the uncomplexed enzyme (Newman *et al.*, 1991) were aligned with the complexed enzyme using MNYFIT (Sutcliffe *et al.*, 1987), omitting the loop region in both enzymes, to a r.m.s. deviation of 0.7 Å. The same algorithm was then used to align only the loop regions of both the uncomplexed and complexed enzymes.

In the region of the specificity pocket S_1 , the flap (residues 71–81) is rotated by approximately 27° with respect to the uncomplexed enzyme co-ordinates and displaced by a maximum of 4 Å at the tip of the loop, closing down onto the bound inhibitor. There is an associated difference in the neighbouring strand 110–115. The main chain conformational changes in the flap (residues 71–81) are essentially localized to the region 73–78 with Gly76 and Gly78 undergoing the largest changes in (Φ, Ψ) of (157°, 126°) and (179°, 22°) respectively. There are few significant differences within the pocket, although the ring of Phe112 moves in towards the cyclohexyl ring of P_1 by 1 Å to give closer packing. The water molecule, which in the native structure sits in the plane of the two catalytic aspartates, is displaced by the hydroxyl group of P_1 which makes an almost identical network of hydrogen bonds with surrounding main chain and side chain atoms.

The flap movement of 71–81 allows more favourable contacts between Gly76, Thr77 and P_2 . Ser219 undergoes a change in its sidechain orientation such that the $O\gamma$ is brought 0.5 Å closer to the $S\gamma$ of P_2 , although the distance is too great for a hydrogen bond; this change in orientation may be due to either a long-range electrostatic interaction or local changes in side chain orientation due to the presence of the bound inhibitor. The hydrogen bond from Ser219 $O\gamma$ to the P_3 peptide nitrogen, seen in other aspartic proteinase inhibitor complexes (Dhanaraj *et al.*, 1992), is retained.

The terminal methyl group of P_2 occupies the same position in space as that of a well-ordered water molecule in the uncomplexed enzyme. The displacement of this water molecule allows the $C\gamma_2$ of Ile30 to move in towards the CH_3 moiety of P_2 by 0.5 Å, providing a more favourable van der Waals interaction. A movement of 1 Å for Gln281 may be a consequence of inhibitor–protein interactions in the region around P_2 but, alternatively, it may be a result of a conformational change in the region 289→297 due to crystal contacts in this region.

The flap provides good van der Waals contacts for the P_2 / P_3 peptide bond, although only Thr77 is close enough to form hydrogen bonds with the inhibitor. The most noticeable difference is the reorientation of the side chain of Ser12, which brings the $O\gamma$ 2 Å closer to the iodine of P_3 . Val111, which makes a large contribution both to this pocket and to P_1 , is displaced by 1.5 Å to bring the carboxylate oxygen closer to the iodine and Gln13 shifts by about 0.5 Å to bring the Ne closer to the iodine. These two differences contribute to the

main electrostatic interactions around the iodine of P_3 , with the remainder provided by the N of Tyr114. Ala115 and Phe117 are displaced by 1 and 0.5 Å respectively, away from P_3 , providing space for the large P_3 residue. The shift of residues Ser12, as mentioned above, of Gln287 towards P_2 and Lys220 provides more space for the P_4 residue in this spacious pocket.

Clearly the most important conformational differences between native and complexed chymosin come from the repositioning of the flap region over the active site. In the uncomplexed crystal form the flap region is involved in intermolecular contacts, possibly stabilizing the observed conformation. In the uncomplexed crystal form the tip of the flap, Thr77, packs onto Ala147 (Thr77 $C\beta$ -Val1 $C\gamma_1$: 3.1 Å) and Val1 (Thr77 $C\gamma_1$ -Ala147 $C\beta$: 3.0 Å) of symmetry related molecules. The contacts between Thr77 and hydrophobic residues of crystallographically related molecules provide stabilization to Thr77 in this position that is not available in solution. Ser79 $C\beta$ is within 3.6 Å of Pro172 $C\gamma$ and also provides good van der Waals contacts between the flap and symmetry related molecules. In the crystals of the inhibitor complex the flap region is far from any intermolecular interactions.

Comparison with a renin–inhibitor complex

The inhibitor in the chymosin–inhibitor complex adopts an extended conformation similar to that found in other aspartic proteinase inhibitor complexes. An alignment of CP-113972 is given with a variety of other aspartic proteinase inhibitors in Figure 1d.

There are two renin–inhibitor complexes available in the Protein Data Bank; 1smr (Dhanaraj *et al.*, 1992; Dealwis *et al.*, 1994), a complex of mouse submaxillary renin with a decapeptide portion of angiotensin (CH-66) and 1rne (Raheul *et al.*, 1991; Dhanaraj *et al.*, 1992), a complex of human recombinant renin with CGP 38'560, an artificial substrate analogue. Although renin and chymosin share only 38% sequence identity, both 1smr and 1rne superpose well with the chymosin–inhibitor complex. With a cut-off of 3.5 Å for structurally equivalent residues, the program MNYFIT (Sutcliffe *et al.*, 1987) found 261 and 264 equivalent residues with chymosin from 1smr and 1rne, respectively, displaying a r.m.s. $C\alpha$ difference of 1.0 Å in each case. However the two renin structures differ with respect to chymosin in the loop regions 246–252 and 156–161, which could be attributed to differing crystal environments, and an insertion in the loop region at 288–292. Because of the similarity between the two renin structures, we have only made detailed comparisons between the chymosin–inhibitor complex and the human recombinant renin–inhibitor complex.

The loop regions 71–81 are found in the same conformation, closed down upon the bound inhibitor, with Tyr75 making a hydrogen bond to Trp39, enclosing the cyclohexyl ring of P_1 . In both complexes a hydrogen bond from the main chain nitrogen of 76 (Ser in renin, Gly in chymosin) to an oxygen atom in the inhibitor further stabilize the loop. The remaining differences in the S_1 pocket are the replacement of Ile30 with Val and Val120 with Ile; the residues differ only by a $-CH_2-$ group and are both hydrophobic. In the chymosin inhibitor complex, Asp215($O\delta_1$) and Gly34(O) pick up additional hydrogen bonds (2.9 Å, 3.4 Å) with the inhibitor P_1 main-chain oxygen (part of the blocking group); this oxygen is not present in the renin inhibitor CGP 38'560. In both models the

rest of this pocket is made up of mainchain contributions from 32→35, 217→219 and hydrophobic contributions from Tyr75, Phe112 and Phe117.

The S₂ pockets are again similar, although some sequence differences give rise to a slightly larger pocket in renin, with alanines replacing Ile300 and Thr218, maintaining hydrophobic contacts, and Ser replacing Val222. In the human renin inhibitor complex the P₂ histidine Nε1 makes a weak hydrogen bond to the Oγ of Ser222 (3.2 Å). The mutation Ser to Val makes this pocket more suitable for the Met-like P₂ SMC of the chymosin inhibitor. The remainder of the pocket is made from mainchain contributions from 213→218. The 288→292 loop in chymosin is a shorter version of the equivalent loop in renin, in which it folds over the active site along with the flap.

S₃ is almost identical in the two enzyme species, the only differences being that renin has a Leu for Tyr114 and a Pro for Val111; both substitutions conserving the hydrophobic nature of the pocket. As with the chymosin complex the renin complex S₄ pocket is very open and the changes Tyr220Lys and His287Gln (renin/chymosin) can make little difference to this pocket, even allowing for the changes in charge and hydrophobicity. The inhibitor in the chymosin complex does not extend far onto the prime side of the scissile bond, where there are significant differences in the specificity pockets.

Some of the reasons for the primary specificity of chymosin for the exposed Phe105–Met106 bond of κ-casein are revealed in the crystal structure of the inhibitor complex. The norstatine residue at P₁ is a good substitute for the phenylalanine of the substrate. The residues comprising S₁ provide tight packing around the cyclohexyl ring at P₁. The same arrangement of side chains in S₁ can be expected around the phenylalanine of the substrate. On the other side of the scissile bond the O_iC₃H₇ moiety of the inhibitor, which partially occupies S₁' makes electrostatic contacts with the enzyme that cannot mimic interactions made by the substrate, which is a methionine at P₁'. On the non-prime side of the scissile bond the inhibitor residues of SMC-pI-Phe-Pro can have few sidechain interactions in common with the substrate sequence Ser-Leu-His. On either end of the substrate binding cleft there are groups of acidic sidechains which may help to align the substrate within the active site cleft, by interacting with the basic residues more than four positions from the scissile bond (N :His-Pro-His-Pro-His-Leu-Ser-Phe-|-Met-Ala-Ile-Pro-Pro-Lys-Lys-Asn: C).

Conclusion

The crystal structure of the complex of chymosin with a reduced bond inhibitor strongly supports the view that the transition state complex of chymosin with its substrate closely resembles that of other aspartic proteinases such as human renin. This is important for the design of engineered chymosins with differing specificities (for example Strop *et al.*, 1990; Nugent *et al.*, 1996; Williams *et al.*, 1997). In these experiments either single residue substitutions or loop replacements were engineered on the basis of the crystal structure of uncomplexed chymosin and modelled inhibitors. The crystal structure of the complex, described here, provides a firmer basis for predicting the effects of these mutations and making more informed designs.

The question of the conformation of the uncomplexed state remains open, although the fact that the flap conformation found in the orthorhombic form makes crystal contacts is indicative of stabilization by intermolecular contacts. However, the alternative possibility is that the conformation observed in

the orthorhombic crystals is also found as a major component in solution (Andreeva, N., personal communication). As a consequence the specificity of chymosin to κ-casein may be linked to a conformational change of the flap, triggered by interactions between chymosin and κ-casein far from the catalytic residues as suggested by Gustchina *et al.* (1996).

References

- Andreeva, N., Dill, J. and Gilliland, G.L. (1992) *Biochem. Biophys. Res. Commun.*, **184**, 1074–1081.
- Badasso, M., Frazao, C., Sibanda, B.L. *et al.* (1992) *J. Mol. Biol.*, **223**, 447–488.
- Blundell, T.L., Jenkins, J.A., Sewell, B.T., Pearl, L.H., Cooper, J.B., Tickle, I.J., Verrapandian, B. and Wood, S.P. (1990) *J. Mol. Biol.*, **211**, 919–941.
- Collaborative Computational Project, Number 4 (1994) *Acta. Crystallogr.*, **D50**, 760–763.
- Chen, L.Q., Erickson, S.W., Rydel, T.J., Park, C.H., Neidhart, D., Luly, J. and Abad-Zapatero, C. (1992) *Acta. Crystallogr.*, **B48**, 476–488.
- Cooper, J.B., Khan, G., Taylor, G., Tickle, I.J. and Blundell, T.L. (1990) *J. Mol. Biol.*, **214**, 199–222.
- Dealwis, C.G., Frazao, C., Badasso, M. *et al.* (1994) *J. Mol. Biol.*, **236**, 342–360.
- Dhanaraj, V., Dealwis, C.G., Frazao, C. *et al.* (1992) *Nature*, **357**, 466–472.
- Dreissen, H., Haneef, I., Harris, G.W. and Howlin, B. (1989) *J. Appl. Crystallogr.*, **2**, 510–516.
- Foltmann, B. (1970) *Methods Enzymol.*, **19**, 421–436.
- Gilliland, G.L., Winborne, E.L., Nachman, J. and Wlodawer, A. (1990) *Proteins*, **8**, 82–101.
- Gustchina, E., Rumsh, L., Ginodman, L., Mujer, P. and Andreeva, N. (1996) *FEBS Lett.*, **379**, 60–62.
- Hoover, D.J., Rosati, R.L. and Wester, R.T. (1989) US Patent 4814342, Example 92.
- James, M.N.G., Sielecki, A.R., Salituro, F., Rich, D.H. and Hoffman, T. (1982) *Proc. Natl Acad. Sci. USA*, **79**, 6137–6141.
- Jollès, J., Alias, C. and Jollès, P. (1968) *Biochem. Biophys. Acta.*, **168**, 591–593.
- Jones, T.A. (1985) *Methods Enzymol.*, **115**, 157–171.
- Leslie, A.G.W. (1993) In Sawyer, L., Isaac, N. and Bailey, S. (eds), *Data Collection and Processing*. Proceedings of the CCP4 Study Weekend, pp. 44–51.
- Lunney, E.A., Hamilton, H.W., Hodges, J.C. *et al.* (1993) *J. Med. Chem.*, **36**, 3809–3820.
- McPherson, A. (1982) *Preparation and Analysis of Protein Crystals*. Wiley & Sons, New York.
- Miller, M., Schneider, J., Sathyanarayana, B.K., Toth, M.V., Marshall, G.R., Clawson, L., Selk, L., Kent, S.B.H. and Wlodawer, A. (1989) *Science*, **246**, 1149–1152.
- Navaza, J. (1994) *Acta. Crystallogr.*, **A50**, 157–163.
- Newman, M., Saffro, M., Frazao, C., Khan, G., Zdobav, G., Tickle, I.J., Blundell, T.L. and Andreeva, N. (1991) *J. Mol. Biol.*, **221**, 1295–1309.
- Nugent, P., Albert, A., Oprayon, P., Blundell, T.L. and Dhanaraj, V. (1996) *Protein Engng.*, **9**, 885–893.
- Parris, K.D., Hoover, D.J., Damond, D. and Davies, D.R. (1992) *Biochemistry*, **31**, 8125–8141.
- Pearl, L.H. and Blundell, T.L. (1984) *FEBS Lett.*, **174**, 96–101.
- Pedersen, V.B. (1977) *Acta. Chem. Scand.*, **B31**, 149–156.
- Pedersen, V.B. and Foltmann, B. (1975) *Eur. J. Biochem.*, **55**, 95–103.
- Priestle, J.P., Fassler, A., Rosel, J., Tintinotblomley, M., Strop, P. and Grutter, M.G. (1995) *Structure*, **3**, 629.
- Raap, J., Kerling, K.E.T., Vreeman, H.J. and Visser, S. (1983) *Arch. Biochem. Biophys.*, **221**, 117–124.
- Raheul, J., Priestle, J.P. and Grutter, M.G. (1991) *J. Struct. Biol.*, **107**, 227–236.
- Säli, A., Veerapandian, B., Cooper, J.B., Foundling, S.I., Hoover, D.J. and Blundell, T.L. (1989) *EMBO J.*, **8**, 2179–2188.
- Säli, A., Veerapandian, B., Cooper, J.B., Moss, D.S., Hofmann, T. and Blundell, T.L. (1992) *Proteins Struct. Funct. Genet.*, **12**, 158–170.
- Schechter, I. and Berger, A. (1967) *Biochem. Biophys. Res. Commun.*, **27**, 157–162.
- Sielecki, A.R., Hayakawa, K., Fujinaga, M., Murphy, M.E.P., Fraser, M., Muir, A.K., Carilli, C.T., Lewicki, J.A., Baxter, J.D. and James, M.N.G. (1989) *Science*, **243**, 1346–1357.
- Sim, G.A. (1959) *Acta. Crystallogr.*, **12**, 813–815.
- Strop, P., Sedalek, J., Stys, J. *et al.* (1990) *Biochemistry*, **26**, 9863–9871.
- Suguna, K., Bott, R.R., Padlan, E.A., Subramanian, E., Sheriff, S., Cohen, G.H. and Davies, D.R. (1987) *J. Mol. Biol.*, **196**, 877–900.
- Suguna, K., Padlan, E.A., Bott, R., Boger, J., Parris, K.D. and Davies, D.R. (1992) *Proteins Struct. Funct. Genet.*, **13**, 195–205.

- Sutcliffe,M.J., Hannef,I., Carney,D. and Blundell,T.L. (1987) *Protein Engng*, **1**, 377–384.
- Tang,J. (ed.) (1977), In *Acid Proteases: Structure, Function and Biology*. Advances in Experimental Medicine and Biology, Vol. 95. Plenum Press, New York.
- Tong,L., Pav,S., Lamarre,D., Pilote,L., LaPlante,S., Anderson,P.C. and Jung,G. (1995) *J. Mol. Biol.*, **250**, 211–222.
- Veerapandian,B., Cooper,J.B., Sali,A. and Blundell,T.L. (1990) *J. Mol. Biol.*, **216**, 1017–1029.
- Williams,M.G., Wilsher,J., Nugent,P. *et al.* (1997) *Protein Engng*, **10**, 991–997.

Received February 14, 1997; revised May 8, 1998; accepted May 18, 1998