Protein Engineering vol.9 no.10 pp.885–893, 1996

Protein engineering loops in aspartic proteinases: site-directed mutagenesis, biochemical characterization and X-ray analysis of chymosin with a replaced loop from rhizopuspepsin

P.G.Nugent, A.Albert, P.Orprayoon, J.Wilsher, J.E.Pitts, T.L.Blundell[1] and V.Dhanaraj

Department of Crystallography, Birkbeck College, University of London, Malet Street, London WC1E 7HX, UK

[1]To whom correspondence should be addressed

The loop exchange mutant chymosin 155–164 rhizopuspepsin was expressed in *Trichoderma reesei* and exported into the medium to yield a correctly folded and active product. The biochemical characterization and crystal structure determination at 2.5 Å resolution confirm that the mutant enzyme adopts a native fold. However, the conformation of the mutated loop is unlike that in native rhizopuspepsin and involves the chelation of a water molecule in the loop. Kinetic analysis using two synthetic peptide substrates (six and 15 residues long) and the natural substrate, milk, revealed a reduction in the activity of the mutant enzyme with respect to the native when acting on both the long peptide substrate and milk. This may be a consequence of the different charge distribution of the mutated loop, its increased size and/or its different conformation.

Keywords: aspartic proteinase/chymosin/crystal structure/mutant

Introduction

Regions joining elements of secondary structure, generally known as loops, play a variety of roles in the structure, evolution, folding and function of proteins (Sibanda and Thornton, 1985, 1991; Sibanda *et al.*, 1989; Greer, 1990). They are often regions providing catalytic or binding residues, essential for function. They are usually the regions where insertions occur in evolution, giving rise to variations in specificity, in thermal stability, in antibody recognition, in proteolytic susceptibility and in protein size, with implications for accessibility to complex oligomeric substrates (Pickersgill *et al.*, 1994). Although the biological consequences of variations in loop regions have often been speculative, the role of loop regions can now be systematically investigated using site-directed mutagenesis. In principle, similar approaches can also be used to engineer novel proteins in which these properties are varied by design.

In this paper, we describe experiments to investigate the role of loops in a family of enzymes, the aspartic proteinases, that have been well characterized in terms of three-dimensional structure, catalytic activity and specificity. For example, several aspartic proteinases have been subjected to extensive structural analyses, including the study of precursors, inhibitor complexes and site-specific mutations, providing an excellent database for the development of the principles of biomolecular design (Davies, 1990). They are of particular interest to such a study as they have evolved to carry out a wide range of functions at different pHs and with different specificities in fungi, plants and animals. They have a considerable commercial importance in drug discovery and in food processing.

Aspartic proteinases are bilobal molecules with a structure consisting of β-sheet with very little α-helix. Each lobe contains a domain of similar topology, but there has been considerable variation in the size of the loop regions including simple replacements of residues, small insertions and deletions and even insertions of small helices and strands (Tang *et al.*, 1978; Blundell *et al.*, 1992). Pepsin-like aspartic proteinases are related to the dimeric retroviral proteinases (Pearl and Taylor, 1987; Blundell *et al.*, 1988; Lapatto *et al.*, 1989; Wlodawer *et al.*, 1989) where the subunits are smaller but have similar topology to the domains of the pepsin-like enzymes; comparisons provide information on how a smaller size might be achieved by engineering in this family of proteinases. β-Sheets of differing size (either four or six strands) form a base to the active sites that lie in deep clefts between the domains and define an identical structural relationship between the catalytic residues on the domains in both pepsin-like and retroviral proteinases.

Although loop regions can adopt a wide variety of conformations, it is now apparent that certain conformational and residue patterns occur frequently (Rose *et al.*, 1985; Sibanda and Thornton, 1985; Ring *et al.*, 1992; Efimov, 1993). Knowledge of families of loops in similar supersecondary structures but unrelated proteins (Sibanda *et al.*, 1989) and in homologous proteins (Sibanda and Thornton, 1993) can be used to design insertions or deletions. Alternatively, a database of supersecondary structures containing two elements of secondary structure and a connecting loop region (Sun and Blundell, 1994; L.Donate, S.Rufino, L.Canard and T.L.Blundell, unpublished results) can be used to select sequences that will adopt stable conformers. However, it is well known that loop structures are most reliably constructed from equivalent loops in homologues (Greer, 1990; Sali *et al.*, 1990; Johnson *et al.*, 1994). In this work, we have adopted this approach in our first experiments but will proceed with novel loop sequences in future work.

We have used chymosin for study as the recombinant enzyme is used commercially for cheese processing and is also involved in soya curd and cocoa processing. It has already been the subject of site-directed mutatgenesis (Strop *et al.*,

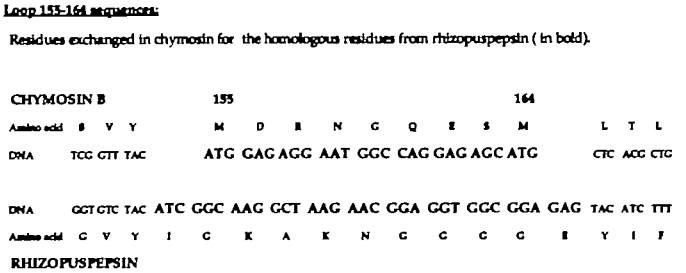


Fig. 1. Sequence comparison of the deleted chymosin loop 155–164 and the corresponding loop from rhizopuspepsin inserted by site-directed mutagenesis.

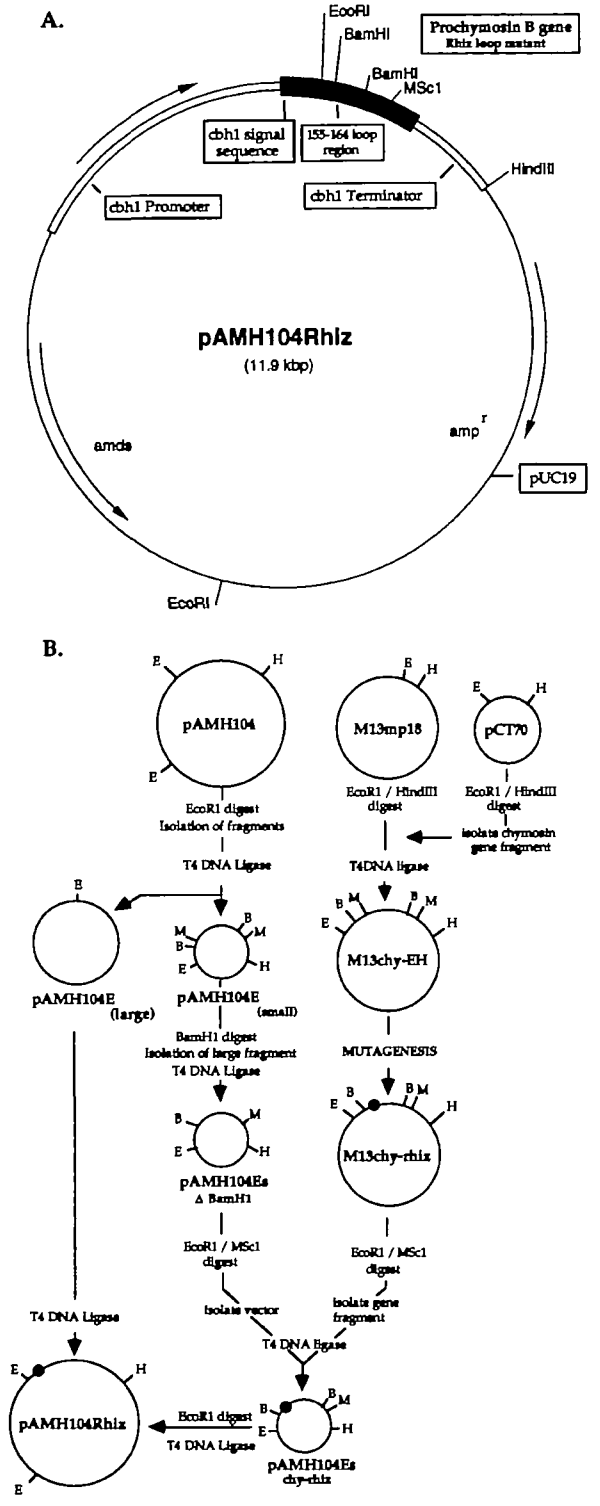
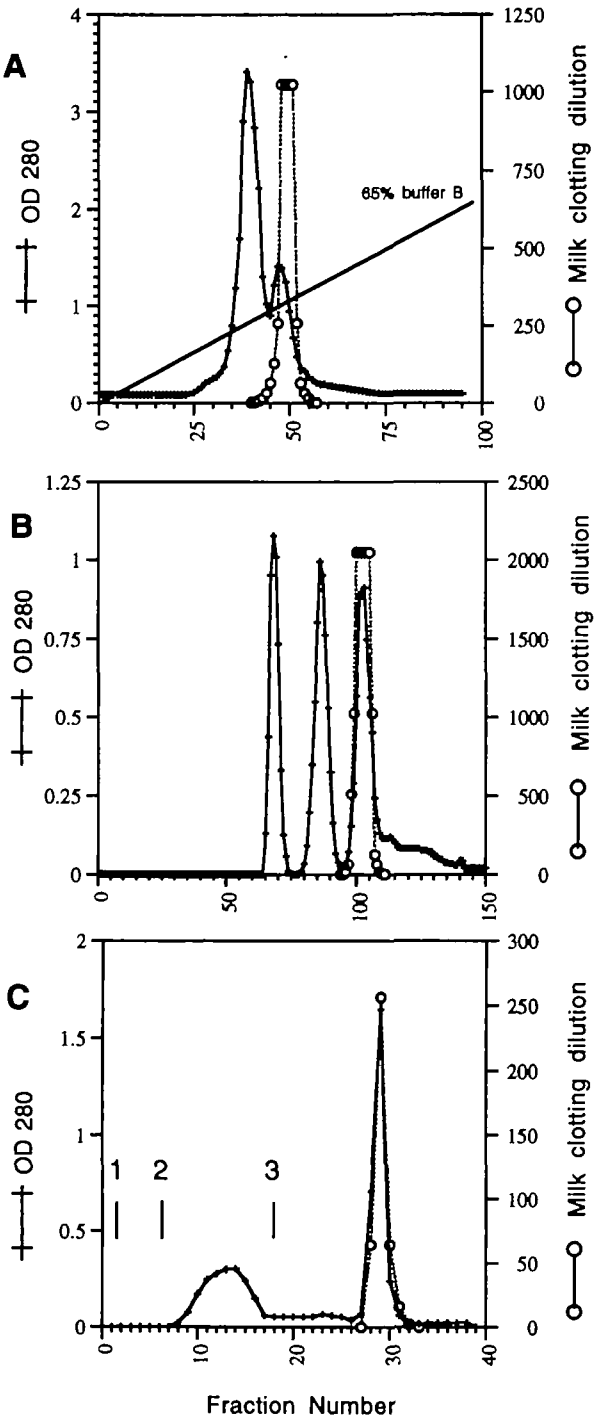


Fig. 2. (A) The chymosin 155–164 rhizopuspepsin expression vector pAMH104Rhiz. (B) Construction of pAMH104Rhiz. Following mutagenesis of M13 chymosin EH, the *Eco*R1–*MSc*1 fragment carrying the mutated loop sequence was removed and ligated into an *Eco*R1–*MSc*1 deleted version of the cloning vector pAMH104$_s$-delta *Bam*H1, to reconstruct the cloning vector pAMH104E$_s$chy-rhiz. Linearization of this vector with *Eco*R1, and its ligation into pAMH104E$_L$ allowed the reconstruction of the complete expression vector pAMH104Rhiz. Introduction of the rhizopus loop sequence was identified by the loss of an *MSc*1 restriction site in M13chy-rhiz, and confirmed by dideoxy sequencing. Restriction sites: E = *Eco*R1, B = *Bam*H1, M = *MSc*1, H = *Hin*dIII.

1990). By comparing the three-dimensional structure of chymosin (Gilliland *et al.*, 1990; Newman *et al.*, 1991) with those of the other members of the aspartic proteinase family, the 155–164 (pepsin numbering) loop was initially targeted for modification by replacement with homologous sequences from the other members of the family as well as with other modelled loops of variable chain length. This loop connects two strands of the basal sheet. Although it is at the end of the active site cleft, located on the protein surface, it does not present contacts with the inhibitor in any of the structures of the aspartic proteinase inhibitor complexes (Bailey and Cooper, 1994). This loop is the most variable in length amongst the aspartic proteinase family.



In this paper, we describe the expression, purification, crystallization, kinetic and structural analysis of such a loop-exchange mutant where the target loop was substituted by the corresponding loop from *Rhizopus chinensis* pepsin.

Materials and methods

Mutagenesis and vector construction

Substitution of the 155–164 loop coding sequence in the calf chymosin B gene by the coding sequence from the corresponding loop from rhizopuspepsin was performed by site-directed mutagenesis using the Oligonucleotide-directed Mutagenesis System (Amersham International) based on the method of Eckstein and co-workers (Sayers *et al.*, 1988). A suitable mutagenesis vector was constructed by the insertion of the *Eco*R1–*Hin*dIII fragment from the *Escherichia coli* metprochymosin expression vector pCT70 (Emtage *et al.*, 1983) into the bacteriophage M13mp18, to construct M13 chymosin EH. This region of pCT70 codes from amino acid residue 118 of mature chymosin.

The mutagenic oligonucleotide 5′ CCCCCAGCGTGAG-CTCTCCGCCACCTCCGTTCTTAGCCTTGCCGATGTAAA-CCGAGAACAG 3′ (mutagenic region underlined) was synthesized using β-cyanoethyl phosphoramidite chemistry (0.2 μmol scale) on an ABI 381A DNA synthesizer, and purified by reversed-phase chromatography on a ProRPC column developed with a gradient of acetonitrile. The nature of the mutation introduced into chymosin B is shown in Figure 1, and was confirmed by dideoxy sequencing.

DNA manipulations for the construction of the expression vector employed (Figure 2) were performed according to standard methodologies (Sambrook *et al.*, 1989). Expression of the prochymosin B gene in this construct is under the control of the strong cellobiohydrolase I (*cbh*1) promoter and the *cbh*1 terminator, employs the cbh1 signal peptide, and is induced by growth in medium containing cellulose. Owing to the presence of the acetamidase (*amds*) gene in this vector, fungal transformants are able to utilize acetamide as a sole nitrogen source.

Transformation of Trichoderma reesei and selection of transformants

Protoplasts of *T.reesei* strain Rut C30 were prepared using the method of Penttila *et al.* (1987) and transformed with the

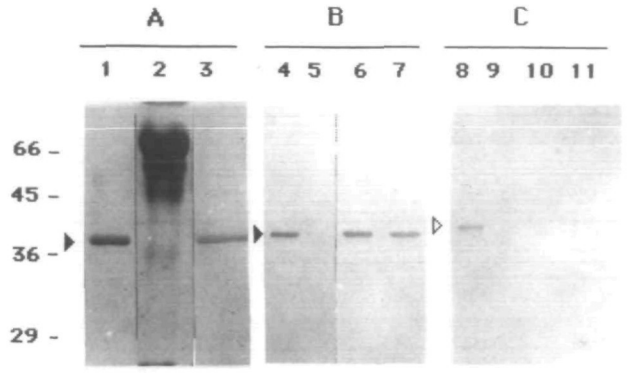


Fig. 4. Analysis of purified chymosin 155–164 rhizopuspepsin by SDS–PAGE and Western blotting. (**A**) SDS–PAGE profiles stained with Coomassie Brilliant Blue. (**B**) Western blotted profiles probed using rabbit anti-chymosin antiserum. (**C**) Western blotted profiles probed using rabbit anti-trichodermapepsin antiserum. Samples applied: standard chymosin B isolated from rennet [lanes 1 (0.5 μg), 4 and 9 (10 ng)], spent fermentation broth [lanes 2 (10 μg), 6 and 10 (25 ng)], affinity purified chymosin mutant enzyme [lanes 3 (0.75 μg), 7 and 11 (20 ng)] and standard trichodermapepsin [lanes 5 and 8 (20 ng)] were analysed. Affinity purified chymosin mutant enzyme was shown to be a single component, entirely free from contamination by trichodermapepsin.

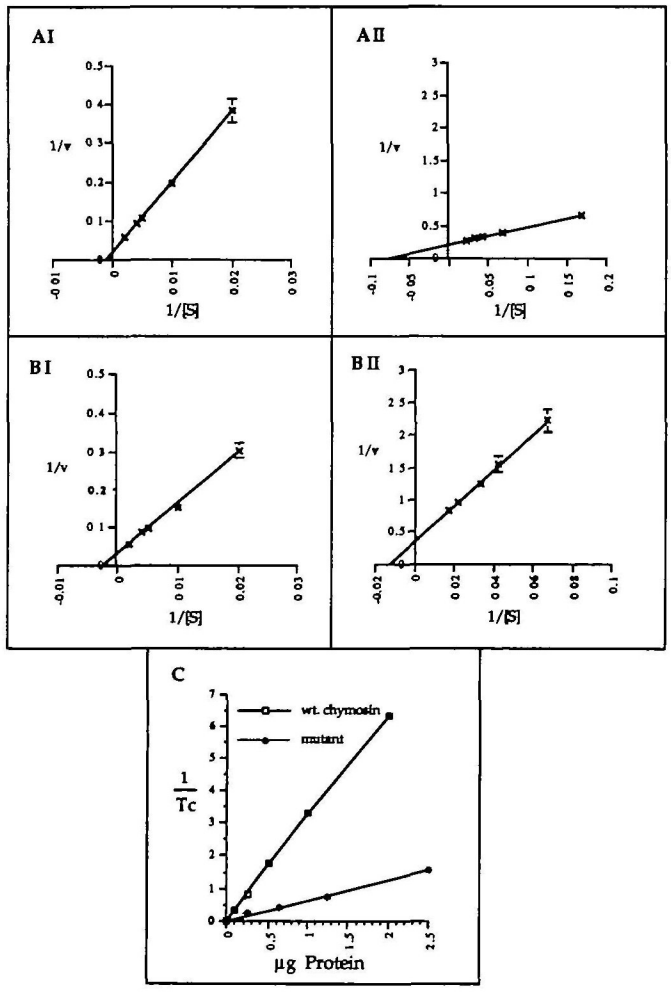


Fig. 5. Typical Lineweaver–Burke plots (**A–C**) for the determination of kinetic parameters for (**A**) standard chymosin B with (**I**) hexapeptide and (**II**) κ-casein 15 residue substrate, and for (**B**) chymosin 155–164 rhiozopuspepsin with (**I**) hexapeptide and (**II**) κ-casein substrate and (**C**) standard curves (turbidimetric) for chymosin B and chymosin 155–164 rhizopuspepsin.

Fig. 3. Purification of chymosin 155–164 rhizopuspepsin. After fermentation, spent culture medium was initially clarified by filtration and centrifugation, and clotting enzyme activity present precipitated by the addition of $(NH_4)_2SO_4$ to 50% saturation. The protein precipitate was resuspended in buffer A (10 mM Tris–10 mM sodium acetate buffer, pH 6.0), and dialysed against the same buffer for 2 days with one change. (**A**) After dilution to 2 l with the same buffer, the enzyme was applied in two batches to a Q Sepharose Fast Flow column (20×5 cm), which was developed with a gradient of buffer B (buffer A + 2 M NaCl). The region of enzyme activity from the two runs was pooled, concentrated to 60 ml in an Amicon stirred cell (PM10 membrane) and (**B**) applied in two batches to an S100 gel filtration column (100×5 cm) developed in 100 mM sodium acetetate buffer, pH 5.6, 150 mM NaCl. The region of enzyme activity from these two runs was again pooled and concentrated as above. Final purification of mutant enzyme was performed by affinity chromatography on a 5 ml Affi-prep 10 column substituted with the peptide inhibitor Val–D-Leu–Pro–Phe–Phe–Val–D-Leu, using a protocol modified from Strop *et al.* (1990). (**C**) Concentrated enzyme from the S100 columns was dialysed against 30 mM sodium formate buffer, pH 4, and (1) applied to the affinity column, (2) unbound protein was washed from the column in the same buffer and (3) active enzyme was eluted in 50 mM sodium phosphate buffer, pH 6.0.

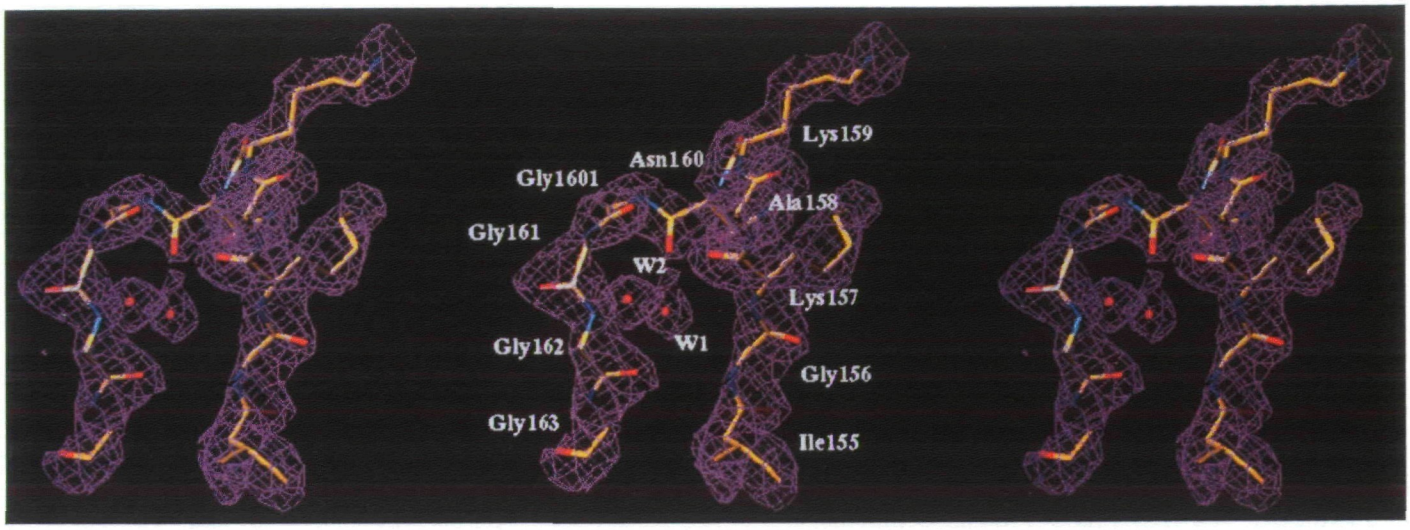


Fig. 6. A section (O; Jones *et al.*, 1991) of the $2|F_o| - |F_c|$ map, contoured at 1σ, showing the structure of the mutated loop.

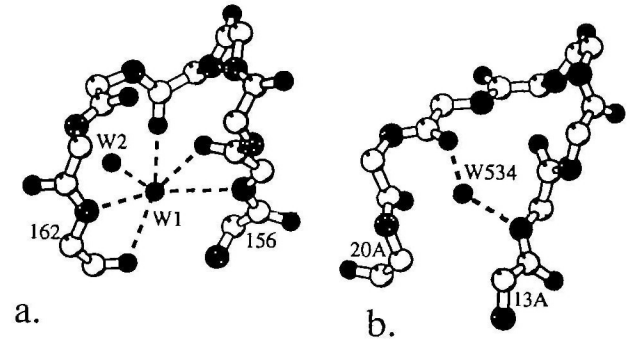


Fig. 7. A view (Molscript and Kraulis, 1991) of (**a**) loop 156–162 of chymosin 155–164 rhizopuspepsin and (**b**) loop 13A–20A of histocompatibility antigen. The hydrogen bonding pattern to chelated water molecules is shown.

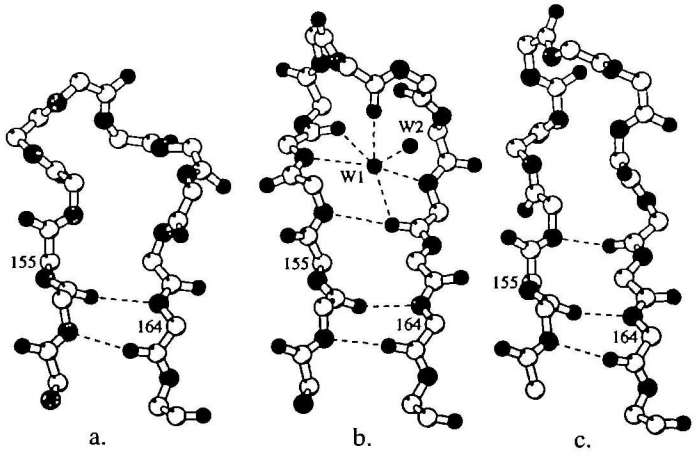


Fig. 8. Main chain conformation and hydrogen bonding pattern of loop 155–164 (**a**) in native chymosin, (**b**) in the loop exchange mutant and (**c**) in rhizopuspepsin (Molscript and Kraulis, 1991).

expression vector pAMH104Rhiz using the PEG–$CaCl_2$ method of Ballance *et al.* (1983). Transformants were selected by three rounds of growth on agar plates containing acetamide as the sole nitrogen source, the first in the presence of sorbitol as an osmotic stabilizer.

Expression and purification of mutant chymosin

Rut C30-Rhiz was grown in 10 l of medium in a 15 l stirred Applikon Fermenter for 5 days, with control of dissolved O_2, pH and temperature, essentially as described by Harkki *et al.* (1989). Typical purification profiles for the chymosin 155–164 rhizopuspepsin mutant are shown in Figure 3. Starting with approximately 8 l of spent fermentation broth containing 19–20 g of total protein, the final affinity purified enzyme (17 mg; 3.5 ml) represented a yield of ~19%, with a purification factor of ~200. The identity of purified protein was confirmed by Western blot analysis following SDS–PAGE, using rabbit anti-chymosin B antiserum, rabbit anti-trichodermapepsin antiserum and alkaline phosphatase-conjugated goat anti-rabbit IgG (Figure 4).

Purification of standard chymosin

For comparison of kinetic parameters, chymosin B was purified from commercial rennet essentially as described by Foltmann (1970) and then further purified by affinity chromatography as described in the legend to Figure 3.

Milk clotting assays for the detection of enzyme activity

Microtitre plate assay. The rapid identification of chymosin-containing fractions during purification was performed using a microtitre plate-based milk clotting assay, modified from Emtage *et al.* (1983). Serial dilutions of sample (20 µl) were incubated in plate wells with 12% w/v dried milk in 0.2 M sodium acetate, pH 5.4, 10 mM $CaCl_2$, (80 µl) for 1 h at 37°C. The plate was then inverted and dropped a short distance to the bench top. The presence of milk clotting activity was indicated by the formation of a curd at the bottom of a well. The amount of enzyme in a given fraction was estimated from the final dilution of each sample that still formed a firm curd.

Turbidimetric assay. The estimation of relative enzyme specific activity using milk as substrate was performed using a modification of the method of McPhile (1976). Sample (5–50 µl) containing up to 2 µg of enzyme protein was added to a solution (1 ml; 0.5% w/v) of dried skimmed milk in 0.2 M sodium acetate buffer, pH 5.4, 10mM $CaCl_2$ at 30°C. Milk clotting resulting from the enzymatic cleavage of κ-casein was monitored by changing turbidity at 550 nm. A standard curve of $1/T_c$ versus the amount of protein added (Brown and

Table I. Kinetic analysis of chymosin (CMS) and chymosin 155–164 rhizopuspepsin (CHRZ)

Kinetic parameters using hexapeptide substrate

Enzyme	Plot	K_m	V_{max}[a]	k_{cat}	k_{cat}/K_m	Specific activity
CMS	LW	880(1)	49.9(0.1)	14.7(–)	0.017(–)	69.3(0.1)
	DP	685(7)	39.3(0.6)	11.5(0.4)	0.017(–)	55(2)
CHRZ	LW	580(12)	43(4)	13(1)	0.021(0.001)	59(5)
	DP	445(64)	35(5)	10(2)	0.023(–)	50(8)

Kinetic parameters using κ-casein peptide substrate

Enzyme	Plot	K_m	V_{max}[b]	k_{cat}	k_{cat}/K_m	Specific activity
CMS	LW	11.5(0.5)	4.5(–)	104(–)	9.1(0.4)	90(–)
	DP	10.9(0.1)	4.5(–)	103(1)	9.4(–)	90(–)
CHRZ	LW	48(3)	2.2(0.1)	40(–)	0.84(0.06)	40.7(–)
	DP	49(2)	2.4(0.1)	40(–)	0.82(0.04)	40.7(–)

Specific activity using milk substrate

Enzyme	Relative specific activity
CMS	5.3(–)
CHRZ	1.0(–)

LW = Lineweaver–Burke plot ($1/v$ versus $1/[S]$). DP = Dixon plot (v versus $[S]$). Units: K_m (μM); V_{max} (μmol/min), k_{cat} (s^{-1}); k_{cat}/K_m (s^{-1} μmol^{-1}); specific activity (μmol/min/μg). Numbers in parentheses are standard deviations, $n = 3$).
[a]20.0 pmol of enzyme per assay.
[b]0.8 pmol of enzyme per assay.

Table II. φ, ψ and ω angles for the mutated loop (CHRZ), the same loop in chymosin (4CMS) and the same loop in rhizopuspepsin (2APR)

	CHRZ	4CMS	2APR		CHRZ	4CMS	2APR
155	**Ile**	**Met**	**Leu**	**1601**	**Gly**	**Gly**	**Gly**
φ	–122	–100	–116	φ	73	52	105
ψ	146	119	124	ψ	–165	49	–6
ω	–179	175	–175	ω	174	–172	179
156	**Gly**	**Asp**	**Gly**	**161**	**Gly**	**Gln**	**Gly**
φ	–102	–72	–85	φ	–137	–140	75
ψ	–90	121	173	ψ	97	–48	–143
ω	179	171	178	ω	–177	–171	–179
157	**Lys**	**Arg**	**Lys**	**162**	**Gly**	**Glu**	**Gly**
φ	–169	–58	–109	φ	149	–98	78
ψ	163	–47	132	ψ	–146	10	–166
ω	171	–178	179	ω	179	171	–179
158	**Ala**	**Asn**	**Ala**	**163**	**Gly**	**Ser**	**Gly**
φ	–99	–64	–57	φ	176	–41	155
ψ	83	4	–36	ψ	–146	133	170
ω	–179	174	180	ω	–174	–173	180
159	**Lys**	**–**	**Lys**	**164**	**Gly**	**Met**	**Gly**
φ	148	–	–60	φ	–177	–147	–159
ψ	9	–	–29	ψ	136	133	133
ω	177	–	180	ω	179	160	176
160	**Asn**	**–**	**Asn**				
φ	–141	–	–103				
ψ	77	–	21				
ω	–175	–	180				

Collinge, 1986) is a straight line up to 2 μg added enzyme protein (T_c = time in seconds for the turbidity reading to reach 0.2 OD). A unit of chymosin activity was defined as the amount of protein (μg) which would produce an OD change of 0.2 in 1 min.

Enzyme kinetic analysis using synthetic peptide substrates

Kinetic values for chymosin 155–164 rhizopuspepsin mutant enzyme and chymosin B were determined using two synthetic peptide substrates related to the sequence of the natural chymosin substrate κ–casein. Substrate A, Leu–Ser–Phe(NO_2)–Nle–Ala–Leu–OMe (Martin *et al.*, 1980), was purchased from Sigma Chemical. Substrate B, His–Pro–His–Pro–His–Leu–Ser–Phe–Phe(NO_2)–Ala–Ile–Pro–Pro–Lys–Lys, was synthesized on an ABI Model 431A peptide synthesizer using Fmoc chemistry. The substrate was purified on a C_{18} reversed-phase column developed with a gradient of acetonitrile in 0.1% TFA.

For each substrate, initial rate measurements of enzymatic cleavage at the scissile bond were determined on a Perkin-Elmer Lambda 6 spectrophotometer at the optimum pH for

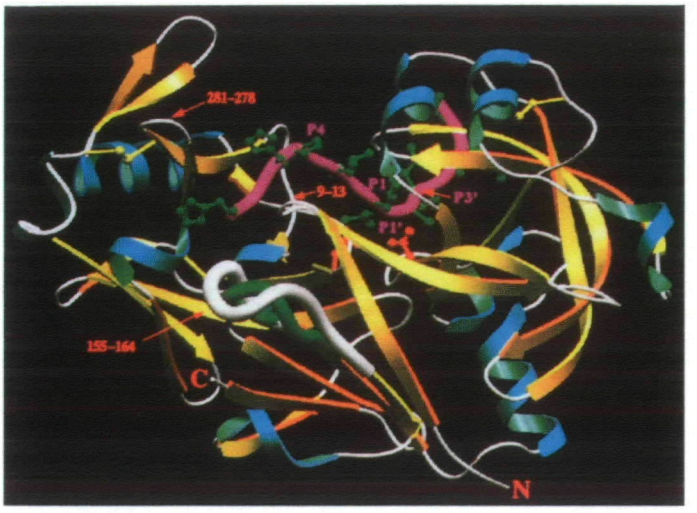


Fig. 9. Model of κ-casein substrate (pink) transition state complex with the loop exchange mutant and with wild-type chymosin (SETOR; Evans, 1993). The figure shows one of the possible conformations of the substrate. The mutated loop is highlighted as a white circular ribbon and the equivalent loop in native chymosin as a green circular ribbon. The catalytic aspartate side chains are displayed in red.

each substrate (A, pH 4.7, 50 mM sodium acetate buffer; B, pH 5.3, 0.1 M sodium acetate, 1 mM $CaCl_2$) at 30°C, by monitoring the change in absorbance at 310 nm (Visser and Rollema, 1986). Kinetic parameters were determined from plots of $1/v$ versus $1/[S]$ and v versus $[S]$. Owing to the lack of a tight-binding, specific inhibitor for chymosin, k_{cat} values were based on an estimate of enzyme concentration using an $E^{1\%}1$ cm 280 nm value of 14 for purified chymosin B (Rothe *et al.*, 1976).

Crystallization and data collection

The purified mutant was concentrated to 10 mg/ml in sodium phosphate buffer at pH 5.6 and crystallized by the hanging drop vapour diffusion method using sodium chloride as the precipitant. Plate-like crystals of dimensions 0.20×0.15×0.05 mm grown over a period of 3–4 weeks were found to be isomorphous with the native crystals.

Intensity data were collected by the oscillation technique with a MAR Research image plate detector using the Daresbury synchrotron radiation source (Station RX 9.5, wavelength 0.92 Å). Although the diffraction extended to 2.2 Å, the images showed splitting and the intensities were weak owing to the small size of the crystals at resolutions beyond 2.5 Å. Oscillation images from two differently oriented crystals were processed using the CCP4 software suite (CCP4, 1994) and merged to obtain a unique set of 14 731 reflections [R_{merge} 14.8%, completeness 92.1%, $I/\sigma(I)$ 27.5, multiplicity 3.1] extending to a resolution of 2.2 Å.

Structure solution and refinement

The atomic coordinates of native chymosin B (Newman *et al.*, 1991), after the removal of the mutated loop, were used to obtain the initial difference Fourier maps. Some cycles of an interactive model-building process, on an Evans and Sutherland PS390 instrument with the graphics software FRODO (Jones, 1978), followed by a least-squares refinement using RESTRAIN (Haneef *et al.*, 1985), gradually revealed the atomic positions of the mutated loop. At this stage, neither the 5–15 nor the 287–300 loop was defined, and this fact together with the high noise level of the $2|F_o|-|F_c|$ map prompted the use of molecular dynamics in subsequent steps of the refinement. Prior to that, some poorly defined residues were replaced by alanines to avoid bias in the refinement. Simulated annealing with XPLOR (Brunger, 1988) allowed the definition of previously ill-defined loops and the location of most of the side chains previously removed. Again, the model building process, followed by a least-squares refinement (XPLOR; Brunger, 1988), resulted in the present model which has 3710 atoms and an R factor of 20.1% between 8 and 2.5 Å resolution level (21.5% between 8 and 2.2 Å). The model exhibits good stereochemistry, with 97.8% of the residues lying in the allowed regions of the Ramachandran plot (PROCHECK; Morris *et al.* 1992). The highest B factors correspond to those atoms belonging to surface loops. The electron density is of good quality for the main chain and most of the side chains, in both the surface loops, including the mutated one, and the core of the enzyme. Gln106 and Met155 side chains presented poor electron density and were replaced by alanines in the last stages of the refinement.

Results and discussion

The loop substitution mutant of bovine chymosin B, in which loop 155–164 was replaced with the equivalent loop from rhizopuspepsin, folded successfully upon expression by *T. reesei*, and was secreted into the medium. Kinetic analysis of the purified mutant enzyme showed that while it was enzymically active, the extent of its enzymatic activity was dependent on the substrate employed, in a manner distinct from that of the wild-type chymosin B (Figure 5 and Table I).

With a hexapeptide substrate, the loop substitution caused no significant change to the substrate binding ability or the catalytic efficiency of the enzyme. However, when a 15 residue (κ-casein peptide) synthetic substrate was employed, the mutation was found to result in a small (twofold) reduction in enzyme catalytic efficiency (k_{cat}), and a much greater reduction in the strength of substrate binding, as indicated by a fivefold increase in the value of K_m for this substrate. The pronounced difference in the activity of the mutant and wild-type enzymes towards these two substrates is very clearly shown by the specificity constant (k_{cat}/K_m) in each case. With the small synthetic substrate, the loop substitution had almost no effect on the specificity constant, whereas with the larger synthetic substrate, the mutation reduced the specificity constant 10-fold.

This dependence of enzymatic activity on the length of substrate employed was further demonstrated when κ-casein in whole milk was used as a substrate. In this case, the mutant enzyme exhibited an even greater reduction (fivefold) in its catalytic efficiency, relative to that of the wild-type enzyme.

Surprisingly, although the enzyme activity of the mutant protein was modified, the structural analysis showed no significant differences between the present model and native chymosin B except in the region of the mutated loop (Newman *et al.*, 1991). All the surface loops have the same orientation, even those adjacent to the mutated loop, and both structures share a similar scheme of inter- and intramolecular contacts. This mutation, therefore, seems not to have affected the overall structure of the protein. The main differences are found in the side chains in the regions that were poorly defined in the native protein (Gilliland *et al.*, 1990; Newman *et al.*, 1991).

The structure of the mutated loop is shown in Figure 6 together with the corresponding section of the $2|F_o| - |F_c|$ map. The most relevant feature is the presence of a central water molecule chelated through many contacts. Figure 7(a) shows

Table III. List of hydrogen bonding pairs made by the mutated loop (CHRZ) and the non-mutated loop in wild-type chymosin (4CMS) and the rest of the protein in comparison with the corresponding interactions in wild-type rhizopuspepsin (2APR)

Interacting residues in CHRZ, 4CMS	CHRZ	4CMS	2APR	Interacting residues in 2APR
	Ile155	**Met155**	**Leu155**	
H_2O	MO,MO	MO,MO	MO,MO,MO	**H_2O**
	Gly156	**Asp156**	**Gly156**	
Asn8	–	–	MS	**Asp8**
Tyr14	–	–	MS	**Tyr14**
	Lys157	**Arg157**	**Lys157**	
Thr270	–	–	SS	**Ser270**
Ser272	–	–	SS	**Asp272**
Glu308	–	SM,SM	SS	**Asn308**
Ile326	–	SS,SS	SS	**Glu326**
H_2O	MO*,MO*	MO,SO,SO,SO	–	**H_2O**
	Ala158	**Asn158**	**Ala158**	
Asn8	–	–	MS	**Asp8**
Asp11	MS	–	MS	**Asp11**
H_2O	MO	–	MO	**H_2O**
	Lys159		**Lys159**	
Ser272	–		SS	**Asp274**
H_2O	–		MO	**H_2O**
	Asn160		**Asn160**	
Ile326	SM		SS	**Glu325**
H_2O	MO,MO*,SO,SO		SO	**H_2O**
	Gly1601	**Gly1601**	**Gly1601**	
	Gly161	**Gln161**	**Gly161**	
H_2O	MO,MO	MO	MO	**H_2O**
	Gly162	**Glu162**	**Gly162**	
H_2O	MO*	–	MO	**H_2O**
	Gly163	**Ser163**	**Gly163**	
Leu6	MM	MM	MM	**Met6**
H_2O	MO,MO	MO	MO	**H_2O**
	Gly164	**Met164**	**Gly164**	
H_2O	–	–	MO,SO,SO,SO	**H_2O**

MM stands for main chain–main chain hydrogen bond, MO for main chain–water molecule, MS for main chain–side chain, SS for side chain– side chain and SM for side chain– main chain. O* indicates that this contact is with the inserted water molecule in the loop; – indicates no interaction.

the hydrogen bonding pattern of the water molecule, which interacts with five loop main-chain atoms and with another water molecule. Water molecules have been observed inserted in α-helices, between β-strands and in β-bulges (Loris *et al.*, 1994; Blaber *et al.*, 1995), but this hydrated structure, involving a loop, has not been observed previously. These water-containing structures have been proposed to be trapped folding intermediates (Sundaralingam and Sekharudu, 1989; Finer-Moore *et al.*, 1992).

Comparisons of the 155–164 hairpin in the mutated protein with the equivalent ones in native chymosin and in rhizopuspepsin are shown in Figure 8. In spite of the hairpin of chymosin being two residues shorter than the rest, all the loops are eight residues in length. Thus, the insertion of additional residues in chymosin is accommodated by extending the hairpin away from the core of the protein and forming an extra hydrogen bond. Differences in the phi–psi conformations of the residues in the loops are more evident in the residues linked to the chelated water molecule (see Table II and Figure 8).

A survey of a database containing 2025 loops from 225 representative protein structures of resolution better than 2.5 Å (L.Donate, S.Rufino, L.Canard and T.L.Blundell, unpublished results) revealed 81 loops of the same length in β-hairpins. Interestingly, the equivalent loop in rhizopuspepsin is the most similar to the mutated loop in terms of the main chain atoms r.m.s. deviation. On the other hand, the equivalent hairpin in chymosin exhibits the same pattern as the mutant in terms of the connecting elements of secondary structure. This analysis revealed that the second closest loop to the mutated one in terms of r.m.s. deviation is the 13A–20A loop in the histocompatibility antigen HLA-B*2705 (Saper *et al.*, 1991) which also contains a water molecule in the middle of the loop. As shown in Figure 7(b), the position of this water is not equivalent to that found in the chymosin mutant as it is not hydrogen bonded at the tip of the loop.

The mainchain–mainchain hydrogen bonding patterns between the 155–165 loop and the adjacent loops have been found to be the same in the compared proteins.

The decrease in the number of polar residues in mutated

chymosin with respect to native chymosin (see Table III) does not produce significant changes in the hydrogen bonding to the loop since in chymosin Asp156, Asn158, Gln161, Glu162 and Ser163 side chains are oriented towards the solvent. This implies that the surface of the mutated protein presents a different charge distribution in the area of the loop. As shown in Table III, differences are found in chymosin at Arg157 which is interacting with Glu308 and Ile326. Mutation does not leave this residue out of interactions since Asn160 of the mutant chymosin is hydrogen bonded with it.

The main differences between the mutated protein and rhizopuspepsin are found in Lys157 and Lys159 side chains, which are oriented towards the solvent in the mutated chymosin, and are hydrogen bonded to Ser270, Asp272, Asn308 and Glu326 (being Thr, Ser, Glu and Ile in mutated chymosin) and to Asp272 (Ser in mutated chymosin), respectively (see Table III) in rhizopuspepsin.

Calculations of the number of contacts of the mutated loop with the rest of the protein for chymosin 155–164 rhizopuspepsin, for the equivalent loop in chymosin and for the loop belonging to rhizopuspepsin modelled into chymosin 155–164 rhizopuspepsin protein (cut-off distance of 4.2 Å) show that the mutated chymosin and the native chymosin present almost the same number of contacts, 79 and 82, respectively, even though chymosin has two residues fewer. The number of contacts in the model is 65, so it seems that the number of contacts on the loop ought to remain constant. Therefore, the loop alone accommodates the modification in the sequence and the addition of the water molecule, rather than the rest of the protein structure.

In order to assess the effect of the mutation on the kinetics of proteolysis of the 15 residue substrate, the chymosin substrate transition-state complex was modelled. The structures of a chymosin inhibitor complex (coordinates kindly provided by M.Groves) and the mouse renin inhibitor complex (Dhanaraj *et al.*, 1992) were least-squares fitted to the chymosin mutant using the MNYFIT suite of programs (Sutcliffe *et al.*, 1987). The coordinates of the flap region of the modelled chymosin were taken from the chymosin complex since, in the native form, Tyr75 is blocking the access to subsite S_1 (nomenclature of Schechter and Berger, 1967). The conformation of the mouse renin inhibitor (Dhanaraj *et al.*, 1992; Dealwis *et al.*, 1994) was used to guide the modelling of the peptide substrate for the seven well-defined substrate subsites, S_4 to $S_{3'}$ since it is the largest available substrate analogue. The conformation of the model was optimized manually (FRODO; Jones, 1978) for steric interactions between enzyme and substrate. It was evident from these studies that some conformations of the substrate are able to interact with the mutated 155–164 loop, as is shown in Figure 9.

The change in the enzymatic activity of the mutated protein, with respect to the native protein, towards the 15-residue κ-casein peptide substrate, and κ-casein in whole milk could be a result of direct interference with large peptides or through charge interaction. The minor alteration found in the solvent structure close to the active site does not seem to be the explanation for this, since the activity of both enzymes was similar towards the short substrate. The different charge distribution of the mutated loop, as is indicated in the analysis of the hydrogen bonding pattern, and/or its increased size, however, may be responsible for this change in activity, either by charge repulsion interfering with the binding of in-coming substrate, or by direct steric hindrance of substrate binding by the mutated loop.

Acknowledgements

We thank Merja Penttila (VTT, Finland) for the gift of *T.reesei* Rut C30 and the plasmid pAMH104. A.A. is recipient of a Biotechnology Program fellowship from the EU. We thank the Thai Government for a studentship for P.O. We are grateful for the financial support from the BBSRC.

References

Bailey,D. and Cooper,J.B. (1994) *Protein Sci.*, **3**, 2129–2143.
Ballance,D.J., Duxton,F.P. and Turner,G. (1983) *Biochem. Biophys. Res. Commun.*, **112**(1), 284–289.
Blaber,M., Baase,W.A., Gaassner,N. and Matthews,B.W. (1995) *J. Mol. Biol.*, **246**, 317–330.
Blundell,T. *et al.* (1988) *Eur J. Biochem.*, **172**, 513–520.
Blundell,T.L , Cooper,J.B., Sali,A. and Zhu,Z.Z. (1992) *Adv. Exp. Med. Biol.*, **306**, 443–453.
Brown,R.J. and Collinge,S.K. (1986) *J. Dairy Res.*, **69**, 956–958.
Brunger,A.T (1988) *X-PLOR (version 3.1) Manual* Yale University, New Haven, CT.
CCP4 (1994) Collaborative Computational Project No.4, *Acta Crystallogr.*, **D50**, 760–763.
Dealwis,C.G. *et al.* (1994) *J. Mol. Biol.*, **236**, 342–360.
Dhanaraj,V. *et al.* (1992) *Nature*, **357**, 466–471.
Davies,D.R. (1990) *Annu. Rev. Biophys. Chem.*, 189–215.
Efimov,A.V. (1993) *Prog. Biophys Mol Biol.*, **60**, 201–239.
Emtage,J.S., Angal,S., Doel,M.T., Harris,T.J R., Jenkins,B., Lilley,G. and Lowe,P.A. (1983) *Proc Natl Acad. Sci. USA*, **30**, 3671–3675.
Evans,S.V. (1993) *J. Mol. Graphics*, **11**, 134–138.
Finer-Moore,J.S., Kossiakoff,A.A., Hurley,J.H , Earnest,T. and Stroud,R.M. (1992) *Proteins: Struct Funct. Genet.*, **12**, 203–222.
Foltmann,B. (1970) *Methods Enzymol.*, **19**, 421–436.
Gilliland,G.L., Winborne,E.L., Nachman,J. and Wlodawer,A. (1990) *Proteins*, **8**, 82–101.
Greer,J. (1990) *Proteins*, **7**, 317–334.
Haneef,I , Moss,D., Stanford,M.J. and Borkakoti,N. (1985) *Acta Crystallogr.*, **A41**, 426–433.
Harkki,A., Uusitalo,J., Bailey,M., Penttila,M and Knowles,J.K.C. (1989) *Biotechnology*, **7**, 596–603.
Johnson,M.S., Srinivasan,N., Sowdhamini,R. and Blundell,T.L. (1994) *Crit. Rev. Biochem. Mol. Biol.*, **29**, 1–68.
Jones,T A (1978) *J Appl. Crystallogr.*, **1**, 268–272.
Jones,T.A., Zou,J.Y., Cowan,S.W and Kjeldaard,M. (1991) *Acta Crystallogr.*, **A47**, 110–119.
Lapatto,P *et al.* (1989) *Nature*, **342**, 299–302.
Loris,R., Stas,P.P.G. and Wyns,L. (1994) *J. Biol. Chem.*, **269**, 26722–26733.
Martin,P., Raymond,M.-N., Bricas,E. and Ribadeau Dumas,B. (1980) *Biochim. Biophys. Acta*, **612**, 410–420
McPhile,P. (1976) *Anal. Biochem.*, **73**, 258–261.
Molscript and Kraulis,P.J. (1991) *J. Appl. Crystallogr.*, **22**, 946–950.
Morris,A.L., MacArthur,M.W., Hutchinson,E.G. and Thornton,J.M. (1992) *Proteins: Struct. Funct Genet.*, **12**, 345–364.
Newman,M., Safro,M., Frazao,C., Khan,G., Zdanov,A., Tickle,I.J., Blundell,T.L. and Andreeva,N. (1991) *J. Mol. Biol.*, **221**, 1295–1309.
Pearl,L.H. and Taylor,W.R. (1987) *Nature*, **329**, 351–354.
Penttila,M., Nevalainen,H., Ratto,M., Salminen,E. and Knowles,J. (1987) *Gene*, **61**, 155–164.
Pickersgill,R. *et al.* (1994) *FEBS Lett.*, **347**, 199–202.
Ring,C.S., Kneller,D.G., Langridge,R. and Cohen,F.E. (1992) *J. Mol. Biol.*, **224**, 685–699.
Rose,G.D., Gierasch,L.M. and Smith,J.S. (1985) *Adv. Protein Chem.*, **37**, 1–109.
Rothe,G.A.L., Axelsen,N.H., Johnk,P. and Foltmann,B. (1976) *J. Dairy Res.*, **43**, 85–95.
Sali,A., Overington,J.P., Johnson,M.S. and Blundell,T.L. (1990) *Trends Biochem. Sci.*, **15**, 235–240.
Sambrook,J., Fritsh,E.F. and Maniatis,T. (1989) *Molecular Cloning. A Laboratory Manual.* 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Saper,M.A., Bjorkman,P.J. and Wiley,D.C. (1991) *J. Mol. Biol.*, **219**, 277–300.
Sayers,J.R., Schmidt,W. and Eckstein,F. (1988) *Nucleic Acids Res.*, **16**, 791–802.
Schechter,I. and Berger,A. (1967) *Biochem. Biophys. Res. Commun.*, **27**, 157–162.

Sibanda,B.L and Thornton,J.M. (1985) *Nature*, **316**, 170–174.
Sibanda,B.L. and Thornton,J.M. (1991) *Methods Enzymol.*, **202**, 59–82.
Sibanda,B.L. and Thornton,J.M. (1993) *J. Mol Biol.*, **229**, 428–447.
Sibanda,B.L., Blundell,T.L and Thornton,J.M. (1989) *J. Mol. Biol.*, **206**, 759–777.
Strop,P. *et al.* (1990) *Biochemistry*, **29**, 9863–9871.
Sun,Z. and Blundell,T.L. (1995) In *Proceedings of the 28th Annual Hawaii International Conference on System Sciences*, in press.
Sundaralingam,M. and Sekharudu,Y.C. (1989) *Science*, **244**, 1333–1337.
Sutcliffe,M.J., Haneef,I., Carney,D. and Blundell,T.L. (1987) *Protein Engng*, **1**, 377–384.
Tang,J., James,M.N.G., Hsu,I N., Jenkins,J.A. and Blundell,T.L. (1978) *Nature*, **271**, 618–621.
Visser,S. and Rollema,H.S. (1986) *Anal. Biochem.*, **153**, 235–241.
Wlodawer,A. *et al.* (1989) *Science*, **245**, 616–621.

Received January 31, 1996, revised May 3, 1996; accepted May 15, 1996