The effect of changing the hydrophobic S1' subsite of thermolysin-like proteases on substrate specificity.

Abstract

The hydrophobic S_1' subsite is one of the major determinants of the substrate specificity of thermolysin and related M4 family proteases. In the thermolysin-like protease (TLP) produced by *B. stearothermophilus* (TLP-ste), the hydrophobic S_1' subsite is mainly formed by Phe130, Phe133, Val139 and Leu202. In the present study, we have examined the effects of replacing Leu202 by smaller (Gly, Ala, Val) and larger (Phe, Tyr) hydrophobic residues. The mutational effects showed that the wild-type S_1' pocket is optimal for binding leucine side chains. Reduction of the size of residue 202 resulted in a higher efficiency towards substrates with Phe in the P_1' position. Rather unexpectedly, the Leu202Phe and Leu202Tyr mutations, which were expected to decrease the size of the S_1' subsite, resulted in a large increase in activity towards dipeptide substrates with Phe in the P_1' position. This is probably due to the fact that 202Phe and 202Tyr adopt a second possible rotamer which opens up the subsite compared to Leu202 and which favours interactions with the substrate. To validate these results we constructed variants of thermolysin with changes in the S_1' subsite. Thermolysin and TLP-ste variants with identical S_1' subsites were higly similar in terms of their preference for Phe versus Leu in the P_1' position.

Introduction

Thermolysin-like proteases (TLPs) are members of the peptidase family M4 (51) of which thermolysin (TLN, EC .3.4.24.27) is the prototype. The amino acid sequences of several TLPs have been determined [see (51), or the Merops data base at http://www.merops.co.uk/merops/famcards/m4.htm], and the three-dimensional structure of TLPs isolated from several bacteria have been solved [*Bacillus thermoproteoliticus* (52), *Bacillus cereus* (54, 54), *Pseudomonas aeruginosa* (55) and *Staphylococcus aureus* (56)]. TLPs consist of an α-helical C-terminal domain and an N-terminal domain mainly consisting of β-strands. The domains are connected by a central α-helix. This helix is located at the bottom of the active site cleft and contains several of the catalytically important residues (Fig. 4.1). Four substrate binding pockets [S_2, S_1, S_1' and S_2'; nomenclature according to Schechter and Berger (174)] have been identified (62). The S_1' subsite is a hydrophobic pocket which is considered a major determinant of substrate specificity (114, 133). In thermolysin and the TLP produced by *B. stearothermophilus* (TLP-ste), the subjects of this study, the S_1' subsite is mainly formed by Phe130, Val139, Leu202 and Phe133 (TLN) or Leu133 (TLP-ste).

Crystallographic (52, 61, 62, 155) and modelling studies (62) of thermolysin have indicated that the S_1' subsite allows efficient binding of a leucine side chain. The notion that the S_1' subsite in thermolysin is not optimal for binding larger residues such as phenylalanine (61, 62) was experimentally confirmed by Izquierdo and Stein (175). These authors showed a clear positive correlation between the size of the P_1' residue and the activity of the enzyme on dipeptide substrates of the 3-(2-furylacryloyl)-L-glycyl-L-Xxx-amide type (FaGXa, where Xxx is a hydrophobic amino acid). Phenylalanine, however, did not conform

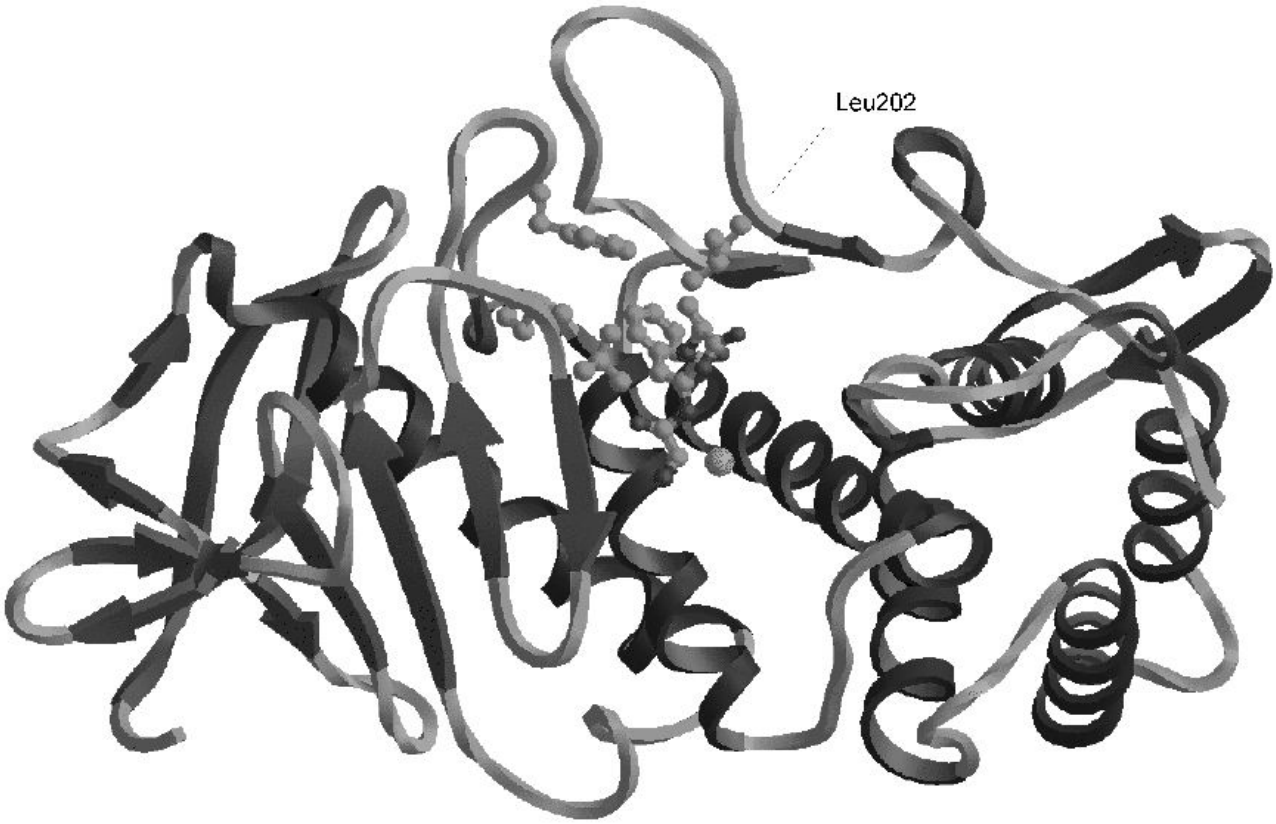

Figure 4.1. Ribbon diagram of thermolysin with a substrate in the active site clef. The left side shows the N-terminal domain consisting predominantly of β-sheet; the right side the predominantly α-helical C-terminal domain. The central α-helix, at the bottom of the active site cleft, contains several residues important for catalysis, including those coordinating the Zn^{2+} ion (blue sphere). The substrate is the tripeptide Gly-Phe-Ala, occupying the S_1 to S_2' positions. The Gly is modeled, whereas the Phe-Ala part is taken from one of the thermolysin structures. The S_1' side chains shown are (counter clockwise, starting at 202Leu); 202Leu, 130Phe, 133Leu and 139Val.

to this trend as illustrated by the fact that similar k_{cat}/K_m values were obtained for FaGLa and FaGFa. The S_1' subsite of TLP-ste is similar in structure and character to that in thermolysin, but TLP-ste has a higher preference for substrates with a Phe at P_1'.

In the present study we have investigated the possibility of modifying the S_1' subsite in TLPs in order to change the preferences of the enzyme for Leu and Phe in the P_1' position. Our hypothesis was that a limited increase in the size of the S_1' pocket could result in an enzyme that would retain its catalytic power, while displaying an increased preference for Phe at position P_1' in the substrate.

As we have shown previously (176), mutating Phe133Leu results in a decreased specificity for P_1' Phe substrates in TLP-ste and changes its substrate specificity into that of TLN. We therefore chose to mutate residue Leu202 which, together with residue 133 dominates the entrance of the substrate to the S_1' pocket (62, 176). We were particularly interested in TLP-ste since we have previously constructed a highly thermostable variant of this enzyme (84). To validate our models of TLP-ste and the S_1' pocket of the M4 peptidases in general, we also constructed and characterized a number of mutants in thermolysin.

Our results show that it is indeed possible to increase the preference for a P_1' Phe by mutating residue 202. The results also

indicate that the substrate specificity of TLP-ste and TLN will change in a similar manner to mutations at positions 133 and 202. This supports the idea that the specificity of other M4 peptidases can be changed in the same way as in TLP-ste and TLN.

Materials and Methods

Modelling and mutant design.

A three-dimensional model of TLP-ste was built on the basis of the crystal structure of thermolysin, using the molecular modelling program WHAT-IF (140), as has been described elsewhere (78). The high sequence identity between thermolysin and TLP-ste (86%) indicates that the TLP-ste model should be sufficiently reliable for prediction and analysis of the effects of most amino acid substitutions (78, 141). Indeed, the TLP-ste model has been used successfully for the design of various stabilizing mutations (43, 82, 142, 177). Throughout this paper, residues in all TLPs are numbered according to the numbering of corresponding residues in thermolysin.

Molecular biological techniques.

The *nprT* gene encoding the TLP of *Bacillus stearothermophilus* CU21 (136) (TLP-ste) was cloned, subcloned, and expressed as described previously (80). The plasmid pUBTZ2 (171) containing the *nprM* gene encoding thermolysin (135) was obtained from DSM-HSC, Geleen, The Netherlands. Site-directed mutagenesis was performed either by the PCR-based mega-primer method, essentially as described by Sarkar and Sommer (139) or with the QuikChange site-directed mutagenesis kit from Stratagene, La Jolla, USA. The QuickChange procedure basically uses a pair of complementary PCR primers that places the mutation in the middle of the primers. pUC18 containing a subcloned fragment of *nprT*, or pUBTZ2 containing *nprM*, was amplified using *Pyrococcus furiosus* DNA polymerase (Pfu Turbo) and these primers for 18 cycles in a DNA thermal cycler. After digestion of the parental DNA with *Dpn*I, the amplified DNA incorporated with the nucleotide substitution was transformed into *E. coli* XL1-Blue strain. Mutagenic primers were designed such that mutant clones could be recognized by the presence or absence of an endonuclease restriction site (80). The nucleotide sequences coding for the mature part of the proteases were verified by DNA sequence analysis. The mutated fragments of TLP-ste were subsequently cloned into the *Bacillus* expression vector pGE501 (178) containing the TLP-ste gene with a deletion of the previously subcloned fragment.

Production and Characterization of Mutant Enzymes.

Production and purification of the enzymes were performed as described earlier (80) using the *B. subtilis* strain DB117 (Δ*npr*,Δ*apr*)(144). Before determining the kinetic parameters, protease preparations were desalted using pre-packed PD-10 gel filtration columns supplied by Amersham Pharmacia, Uppsala, Sweden. Specific activities of the TLP-ste variants towards casein were determined according to a method adapted from Fujii *et al*. (136): approximately 0.5 μg of protease was incubated in 1 ml of 50 mM 2-amino-2-(hydroxymethyl)-1,3-propanediol (Tris·HCl), pH 7.5, containing 0.8% (wt/vol) casein and 5 mM $CaCl_2$ at 37°C for 1 h. Reactions were quenched by the addition of 1 ml of a solution containing 100 mM tri-chloro-acetic acid (TCA), pH 3.5. One unit of activity was defined as the amount of enzyme activity needed to liberate a quantity of acid-soluble peptide corresponding to an increase in A_{275nm} of 0.001 per minute.

The k_{cat}/K_m and K_m values of the enzymes for furylacryloylated di- and

tripeptides were determined at 37°C, in a thermostated Perkin-Elmer Lambda 11 spectrophotometer. The reaction mixture (1 ml) contained 50 mM Tris, 50mM 4-morpholineethanesulfonic acid (MES), pH 7.0, 5 mM $CaCl_2$, 5% Me_2SO, 0.5% 2-propanol, 0.01% Triton X-100 and 100 μM to 2.5 mM of substrate, and the reaction was followed by measuring the decrease in absorption at 345 nm ($\Delta\varepsilon_{345}$ = -317 $M^{-1}\cdot cm^{-1}$)(179). All substrates were supplied by Bachem AG, Bubendorf, Switzerland. Stock solutions of the furylacryloylated dipeptides 3-(2-furylacryloyl)-L-glycyl-L-leucine-amide (FaGLa) and 3-(2-furylacryloyl)-L-glycyl-L-phenyl-amide (FaGFa), and of the furylacryloylated tripeptides 3-(2-furylacryloyl)-L-glycyl-L-leucine-L-alanine (FaGLA) and 3-(2-furylacryloyl)-L-glycyl-L-phenylalanine-L-leucine (FaGFL) were prepared by dissolving the peptides in Me_2SO. Apparent second order rate constants (k_{cat}/K_m) were determined by varying the enzyme concentrations over a 50-fold range under pseudo-first-order conditions and measuring the initial activity, essentially according to the method described by Feder (179).

The K_i for phosphoramidon (*N*-[α-L-rhamnopyranosyl-oxyhydroxyphosphinyl]-L-leucine-L-tryptophan), Roche Molecular Biochemicals, was determined by a 10 minutes preincubation of a 0.1 nM protease solution with varying concentrations of the inhibitor (10^{-8} to 10^{-3} M), in 50 mM Tris, 50 mM MES, pH 7.0, 5 mM $CaCl_2$, 0.01% Triton X100. Subsequently, enzyme activity was determined using 100μM FaGLA as substrate. K_i 's were calculated by the method described by Hunter and Downs (145).

Results

Mutant design and production of mutant proteins.

Position-specific rotamer searches (180) for the residues to be introduced at position 202 and 133 showed that all new side chains could adopt a favorable rotamer without concomitant introduction of steric overlap in both TLP-ste and thermolysin. Furthermore, the modelling studies indicated that the Leu202Gly, Leu202Ala and Leu202Val mutations would simply increase the size of the pocket by a volume corresponding to approximately four, three and one methyl groups, respectively, whereas Leu202Phe and Leu202Tyr were expected to lead to some reduction in size.

The various mutant TLPs were constructed, produced, purified and characterized as described in *Materials and Methods.* Production and purification yields were similar to those of the wild-type TLP-ste. All variants had similar thermal stabilities and specific activities towards casein (data not shown).

Characterization of mutant proteases.

To examine the enzymatic properties of the mutant TLP-ste enzymes, kinetic parameters for the reaction with available tripeptide substrates as well as the K_i for phosphoramidon were determined (Table IV.I). The k_{cat}/K_m values for FaGLA indicate that leucine is the optimal residue at position 202 for substrates with a Leu at the P_1' position. The results with FaGFL show that the activity towards substrates with a phenylalanine at the P_1' position was increased by replacing Leu202 by a smaller residue. For FaGFL, the Leu202Ala mutant was the most active. It is interesting to note that the Leu202Phe and Leu202Tyr mutations were more deleterious for activity towards FaGLA than for activity towards FaGFL (with the larger Phe at P_1'). The Tyr variant was more active than the Phe variant for both substrates.

Table IV.I. Specificity of *B. stearothermophilus* thermolysin-like protease (TLP-ste) variants for tripeptide substrates.

	FaGLA				FaGFL				phosphoramidon
	k_{cat}/K_m [a,b]	k_{cat}/K_m	k_{cat}	K_m	k_{cat}/K_m [a,b]	k_{cat}/K_m	k_{cat}	K_m	K_i [d]
	$s^{-1}\cdot M^{-1} \times 10^{-3}$	relative	$s^{-1} \times 10^{-2}$	$M \times 10^{3}$	$s^{-1}\cdot M^{-1} \times 10^{-3}$	relative	$s^{-1} \times 10^{-2}$	$M \times 10^{3}$	$nM \times 10^{-2}$
L202G	2.6	4.7%	nd[c]	nd[c]	1.3×10^{3}	153.7%	3.6	0.28	8.0
L202A	9.0	16.1%	nd[c]	nd[c]	2.2×10^{3}	266.0%	3.5	0.16	1.8
L202V	53.9	96.5%	1.3	2.5	1.5×10^{3}	186.4%	2.2	0.14	0.7
TLP-ste	55.8	100.0%	2.0	3.6	8.3×10^{2}	100.0%	2.0	0.24	0.2
L202F	9.8	17.6%	0.58	5.9	4.2×10^{2}	50.4%	2.5	0.60	1.6
L202Y	18.3	32.7%	1.7	9.4	1.2×10^{3}	149.8%	2.0	0.16	42.4

[a]Reaction conditions: 50 mMTris, 50 mM MES, 5 mM $CaCl_2$, 0.01% Triton X100, pH 7.0, 37°C. [b]Standard deviations are less than 15% of the value given.[c]could not be determined. [d]Standard deviations are less than 10% of the value given.

Table IV.II. Specificity of *B. stearothermophilus* thermolysin-like protease (TLP-ste) variants for dipeptide substrates.

	FaGLa		FaGFa		FaGFa
	k_{cat}/K_m [a,b]	k_{cat}/K_m	k_{cat}/K_m [a,b]	k_{cat}/K_m	FaGLa
	$s^{-1}\cdot M^{-1} \times 10^{-3}$	relative	$s^{-1}\cdot M^{-1} \times 10^{-3}$	relative	ratio k_{cat}/K_m
L202G	0.2	10%	3.3	97%	14.71
L202A	0.5	22%	27.8	810%	56.39
L202V	2.2	97%	12.7	370%	5.88
TLP-ste	2.2	100%	3.4	100%	1.55
L202F	0.9	42%	11.8	344%	12.62
L202Y	3.1	141%	56.6	1648%	18.12

[a]Reaction conditions: 50 mMTris, 50 mM MES, 5 mM $CaCl_2$, 0.01% Triton X100, pH 7.0, 37°C.
[b]Standard deviations are less than 15% of the value given.

The solubility of the tripeptide substrates was sufficient to allow determination of the k_{cat} and K_m values for TLP-ste variants. A complication of the use of tripeptides is that they also occupy the S_2' pocket. The S_2' pocket could be affected by mutation of Leu202 (62), meaning that the mutational effects displayed in Table IV.I may in part be due to effects on binding of the P_2' residue in the substrate. The Morihara-Tsuzuki specificity order (115) shows that substrates with a P_2' Leu (as in FaGFL) have a higher affinity and a higher k_{cat}/K_m than substrates with a P_2' Ala (as in FaGLA). Since the available tripeptide substrates do not have the same P_2' residue, interpretation of mutational effects in terms of changed preferences for the P_1' residue is not straightforward. Therefore, additional characterization of the TLP-ste variants was conducted using the dipeptide substrates FaGLa and FaGFa. Because of low solubilities of these substrates, only k_{cat}/K_m values could be determined (Table IV.II). In accordance with the results obtained with tripeptide substrates, replacement of Leu202 with a smaller residue resulted in a decrease in k_{cat}/K_m for FaGLa and an increase in k_{cat}/K_m for FaGFa, thus yielding a drastic overall change in the preferences for leucine and phenylalanine at P_1'. Replacement of Leu202 by Phe or Tyr had only modest effects on activity towards FaGLa, in accordance with observations with the tripeptide FaGLA. Most remarkably, replacement of Leu202 by Phe or Tyr resulted in a spectacular increase in the k_{cat}/K_m for FaGFa.

The high sequence identity between TLP-ste and thermolysin (86 %) ensures that the three-dimensional model of TLP-ste is quite accurate, especially in well-conserved regions such as the active site [*e.g.* see (78, 141, 181) for a discussion]. To verify the presumed similarity of thermolysin and TLP-ste with respect to the S_1' subsite, thermolysin was mutated such as to make its S_1' subsite "identical" to that of some of the TLP-ste variants. The subsite in both enzymes is composed of Phe130, Val139, Leu202 and Phe133 in TLP-ste or Leu133 in thermolysin. As shown in Table IV.III, TLN and TLP-ste variants with identical residues at positions 130, 133, 139, and 202 had virtually similar preferences for substrates with Phe versus Leu at the P_1' position. This adds confidence to the notion that the S_1' subsites of TLN and TLP-ste are structurally similar and that the model of TLP-ste can be used to design mutations in this subsite.

Table IV.III. Comparison of the specificity of thermolysin and TLP-ste mutants towards dipeptide substrates.

	Position		k_{cat}/K_m [a,b]		Phe
	133	202	FaGLa	FaGFa	Leu
			$s^{-1}\cdot M^{-1} \times 10^{-3}$	$s^{-1}\cdot M^{-1} \times 10^{-3}$	ratio k_{cat}/K_m
TLN wt	Leu	Leu	12.3	3.9	0.3
	Phe	Leu	10.8	15.6	1.5
	Leu	Tyr	25.3	230.0	9.1
	Phe	Tyr	13.0	186.4	14.3
	Leu	Leu	17.2	6.3	0.4
TLP-ste wt	Phe	Leu	2.3	3.4	1.5
	Leu	Tyr	--	--	--
	Phe	Tyr	3.1	56.6	18.3

[a]Reaction conditions: 50 mMTris, 50 mM MES, 5 mM $CaCl_2$, 0.01% Triton X100, pH 7.0, 37°C.
[b]Standard deviations are less than 15% of the value given.

Discussion

A comparison of the results obtained for thermolysin and TLP-ste shows that these enzymes have virtually identical substrate preferences (expressed in the ratio's of k_{cat}/K_m) if they contain identical residues at positions 133 and 202. This shows the reliability of the TLP-ste model (at least near the S_1' subsite) and the usefullness of thermolysin crystal structures in discussing the effects of the mutations. The activities of the different enzymes towards each substrate show that thermolysin generally is a more active enzyme than TLP-ste. Although we can only speculate as to the origin of this difference, differences in hinge-bending motions (86, 148) or in active site electrostatics attributable to the 42 dissimilarities in amino acid composition may play a role.

Owing to extensive crystallographic studies by Matthews and co-workers (52, 61, 155), considerable information is available concerning the interaction between thermolysin and a variety of ligands. Superposition of a series of enzyme-ligand complexes shows that the residues making up the S_1' subsite have highly invariant positions (0.15 – 0.20 Å rms.), indicating that they hardly adapt to the P_1' residue in the ligand (Fig. 4.2). Instead, the ligand seems to adapt, thus ensuring that specific ligand-enzyme interactions are preserved, regardless of the type of P_1' residue. Ligands with a leucine chain at P_1' show a prominent preserved interaction involving a Cδ atom on the substrate and a Leu133-Cδ and a Leu202-Cδ in the S_1' pocket. In case of shorter P_1' side chains without Cδ, the Cα atoms of the ligand shift (by up to 1.2 Å). Consequently, a carbon in the P_1' side chain occupies the position which is occupied by one of the Cδ atoms in case P_1' is leucine. Thus, contrary to what might be expected, larger substrates do not penetrate the pocket deeper than smaller substrates. In addition to showing that the Cδ atoms of Leu202 have important interactions with the P_1' side chain, the crystal structures also suggest that these Cδ atoms are a major sterical hindrance for accommodating larger side chains (such as Phe) in the S_1' pocket.

Almost any mutation of Leu202 reduced the activity towards substrates with Leu at P_1',

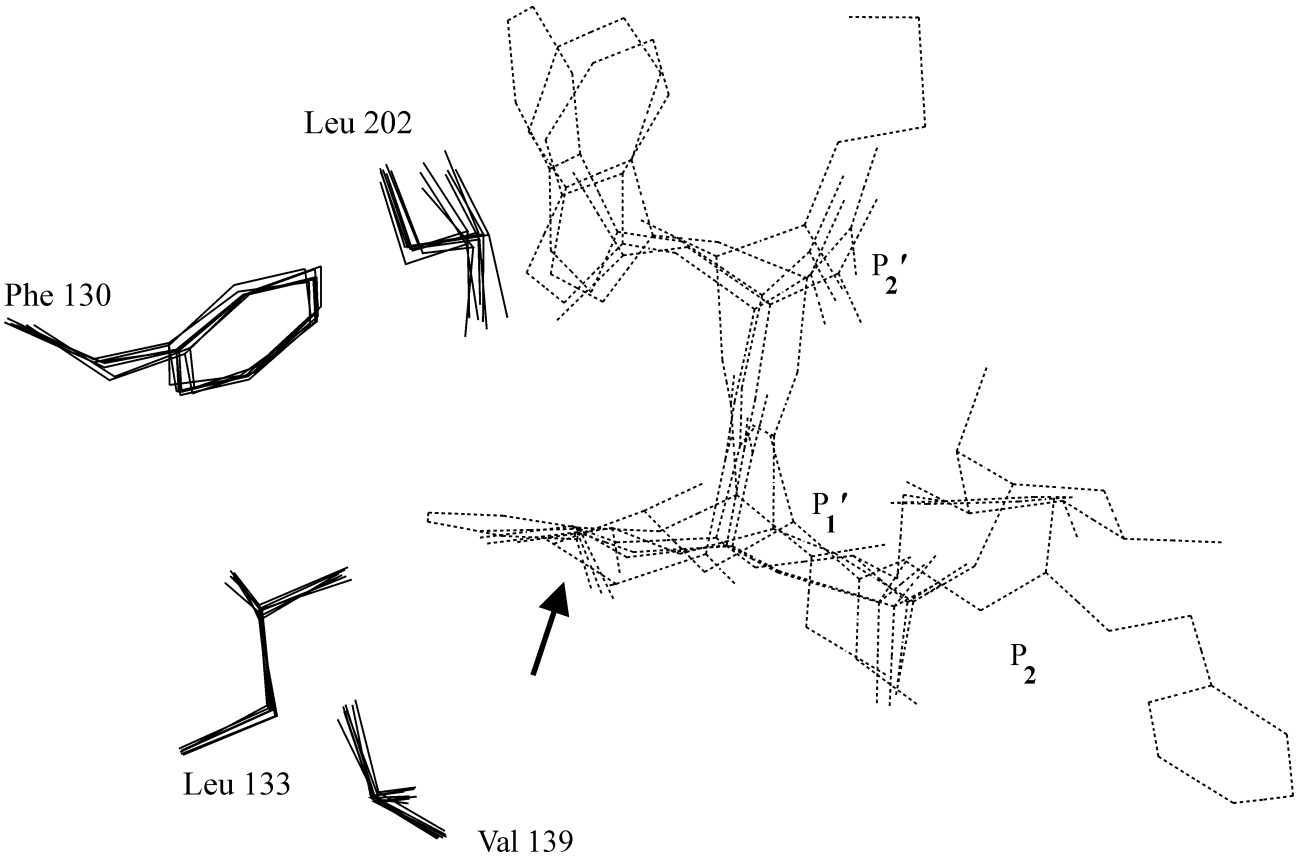

Figure 4.2. The P_1' residues preferentially occupy a specific location in the S_1' subsite. A line diagram showing a superposition of several thermolysin structures (1TLP.PDB, 3TMN.PDB, 2TMN.PDB, 6TMN.PDB, 1TMN.PDB and 4TLN.PDB). The arrow indicates the position that is always occupied by a ligand-carbon atom (see text for details). The superposition shows the remarkable rigidity of the subsite residues and also shows the variability in the positions occupied by the various ligands.

which is in accordance with the notion that in the wild-type enzyme the 202 residue is optimized for binding a Leu in the S_1' pocket. Thus, any other residue at the 202 position will result in lower k_{cat}/K_m values for substrates with a P_1' Leu (62, 175).

Interestingly, it was indeed possible to increase the activity of TLP-ste towards substrates with Phe at P_1' by reducing the size of residue 202. Of the Val, Ala and Gly mutants, the Ala mutant was most active towards substrates with a Phe at P_1'. Modelling studies (not shown) indicated that alanine is the best compromise between on the one hand creating space in the S_1' pocket (not sufficient in Leu202Val) and on the other hand keeping as much as possible contacts with the P_1' side chain (better in Leu202Ala than in Leu202Gly).

If Phe and Tyr would adopt the same rotamer as Leu at position 202, the size and accessibility of the S_1' pocket would be drastically decreased. This would obviously lead to a drastic reduction in catalytic activity towards all substrates tested. Such drastic reductions were not observed, indicating that Phe and Tyr adopt a second favorable side chain conformation in which the aromatic ring is parallel to the substrate. Molecular modelling indicates that the distance between the aromatic rings of 202Phe/Tyr and the P_1' Phe would then be approximately 4 Å, which is appropriate for hydrophobic packing but too large for aromatic stacking.

The Tyr mutant had clearly anomalous characteristics, such as a drastically increased K_i, and higher activity than the Phe mutant for all substrates tested. Most remarkably, the Tyr mutant clearly had the highest activity towards FaGFa of all variants tested in this study. Modelling studies indicate that this anomalous

behavior might be caused by a hydrogen bond between the Tyr-OH and the substrate. A detailed explanation of the remarkable effects of Leu202Tyr awaits crystallographic studies of the mutant.

In case of substrates with a Phe at P_1', mutational effects were more pronounced for the dipeptides than for the tripeptides. This may be due to the fact that residue 202 also affects the S_2' subsite [illustrated by thermolysin-ligand complexes with PDB accession numbers 1TLP, 5TMN and 6TMN (182-184)]. It would not be surprising if effects on the S_2' subsite are most notable for FaGFL, since this substrate has a Leu at P_2'. As explained above, a Leu at P_2' interacts strongly with S_2' (115) and its interactions with the enzyme are thus most likely to be affected by mutation of residue 202.

Several mutagenesis studies of proteases have shown that it is possible to manipulate substrate preferences by changing the size and/or character of hydrophobic binding pockets (95, 96, 98, 113, 185-190). For example, diminishing the space in the S_1 subsite of subtilisin YaB (95) and subtilisin E (185) by increasing the size of the subsite residues led to reduced activity towards substrates with large P_1 residues while yielding higher activity towards substrates with smaller P_1 residues. When it comes to binding of hydrophobic side chains in a subsite, substrate and subsite geometry play a role in addition to subsite size [*e.g.* (96, 113)]. In the present study we probably see both types of effects. The activity effects of reducing the size of residue 202 are likely to be caused at least in part by the increase in size in the S_1' subsite. However, since Leu and Phe are not that different in size and shape, it is likely that mutational effects also reflect changes in the quality of the S_1'-P_1' interactions. Considering the similarity between Leu and Phe, the changes in substrate preferences that were obtained in the present study are remarakbly large, especially for the dipeptide substrates. These changes were obtained by considerable increases in activity for substrates with a P_1' Phe, and not primarily by deterioration of activity towards substrates with a P_1' Leu (in contrast to (186). It is interesting to note that the 16-fold increase in activity that the Leu202Tyr mutant displays towards FaGFa is one of the highest found to date for a single mutant in similar studies.

Acknowledgements

The authors wish to thank Dr Theo Sonke from DSM-Research for supplying pUBTZ2 and helpful discussions.