Engineering high performance variants of Bacillusthermolysin-like proteases

Veltman, Oene Robert

CHAPTER 2

ANALYSIS OF STRUCTURAL DETERMINANTS OF THE STABILITY OF THERMOLYSIN -LIKE PROTEASES BY MOLECULAR MODELLING AND SITE-DIRECTED MUTAGENESIS .

Shortened version published in: *Protein Engineering* **9,** 1181-1190.

SUMMARY

Members of the family of thermolysin-like proteases exhibit large differences in their resistance to thermally induced unfolding and subsequent autolytic degradation. The thermolysin-like protease (TLP) produced by *Bacillus stearothermophilus* CU21 (TLP-ste) differs at 43 positions from the more stable thermolysin (containing 316 residues in total). Because of the high sequence identity between thermolysin and TLP-ste (86%) an accurate three-dimensional model of the latter could be built. The differences between the two enzymes appeared to be scattered over the molecule and they displayed large diversity in terms of the amino acids involved. In 26 cases the differences were analysed by site-directed mutagenesis. Remarkably, considerable destabilising effects were found for some mutations, suggesting that thermolysin can be stabilised even further. A Tyr-rich three residue insertion in TLP-ste (the only deletional / insertional difference between the two enzymes) appeared to make an important contribution to the stability of the enzyme. Mutations with large effects on stability were all localised in the ß-pleated N-terminal domain of TLP-ste, confirming observations that this domain has a lower intrinsic stability than the largely α-helical C-terminal domain. Rigidifying mutations such as Gly58→ Ala, Ala69→ Pro were among the most stabilising ones. Apart from this observation, the analyses did not reveal general 'traffic rules' for stabilising proteins. Instead, the results highlight the importance of context in evaluating the stability effects of mutations.

INTRODUCTION

Several *Bacillus* species secrete a metallo-endopeptidase very similar to the highly stable thermolysin (TLN) produced by *B. thermoproteolyticus*. These so-called thermolysin-like proteases (TLPs; 300-319 residues) contain a zinc ion bound in the active site and a varying number of calcium ions that are important for stability (Matthews *et al.*, 1972, Roche & Voordouw, 1978; Holmes & Matthews, 1982; Pauptit *et al.*, 1988; Priest, 1989; Stark *et al.*, 1992). Large stability differences occur within the family of TLPs (Table 1), making these enzymes an interesting system to study structure-stability relationships (Imanaka *et al.*, 1986; Margarit *et al.*, 1992; Vriend & Eijsink, 1993; Hardy *et al.*, 1994; Eijsink *et al.*, 1995). The amino acid sequences of several TLPs are known (e.g. Takagi *et al.*, 1985; Wetmore *et al.*, 1992; O'Donohue *et al.*, 1994; Figure 1) and the three-dimensional structures of TLN

Table 1. Thermal stability of thermolysin-like proteases from *Bacillus* species.

TLP variant	species	T_{50} (°C)	T_{50} (°C)
Thermolysin	*B. thermoproteolyticus* rokko	86.9	+ 13.5
TLP-cal	*B. caldolyticus* YT-P	81.5	+ 8.1
TLP-ste	*B. stearothermophilus* CU21	73.4	0
TLP-cer	*B. cereus* DSM 3101	68.6	- 4.8
TLP-sub	*B. subtilis* 1A40	58.6	- 14.8

Stability is expressed as T_{50} and as T_{50}, being the stability relative to TLP-ste.

(Matthews *et al*., 1972; Holmes & Matthews, 1982) and the less thermostable TLP of *B. cereus* (Pauptit *et al.*, 1988, Stark *et al*., 1992) have been solved. In TLPs an N-terminal domain can be distinguished that consists mainly of -pleated sheet, and a C-terminal domain that is mainly α-helical. The two domains are connected by a central α-helix (residues 136-152) containing several residues important in catalysis (Figure 2; Matthews *et al*., 1972; Pauptit *et al*., 1988).

At elevated temperatures TLPs become irreversibly inactivated as the result of autolysis. Due to the broad specificity of TLPs (Heinrikson, 1977) conformational features rather than sequence-characteristics of the TLP molecule are thought to determine the sites of autolytic attack (Fontana, 1988). It has been shown that the rate of thermal inactivation is controlled by the rate of local unfolding processes that render the TLP molecule susceptible to autolysis (Dahlquist *et al.*, 1976; Vriend & Eijsink, 1993; Kidokoro *et al.*, 1995; Hardy *et al.*, 1994; see also Fr mmel & Sander, 1989; Braxton & Wells, 1992). The fact that the stability-determining unfolding processes have a local character means that the effect of a site-directed mutation on stability is to a certain extent determined by the location of the mutation: mutations in regions that partially unfold to the most easily autolytically susceptible conformations will have relatively large effects on stability. Such regions are likely to be located at the surface of the protein, since the early steps of unfolding of a protein are thought to involve mainly surface-located structure elements (Matouschek *et al.*, 1989; Jackson & Fersht, 1991; Vriend & Eijsink, 1993). In accordance with this model for thermal inactivation, mutational analyses of TLP-ste have shown that the stability of this enzyme is

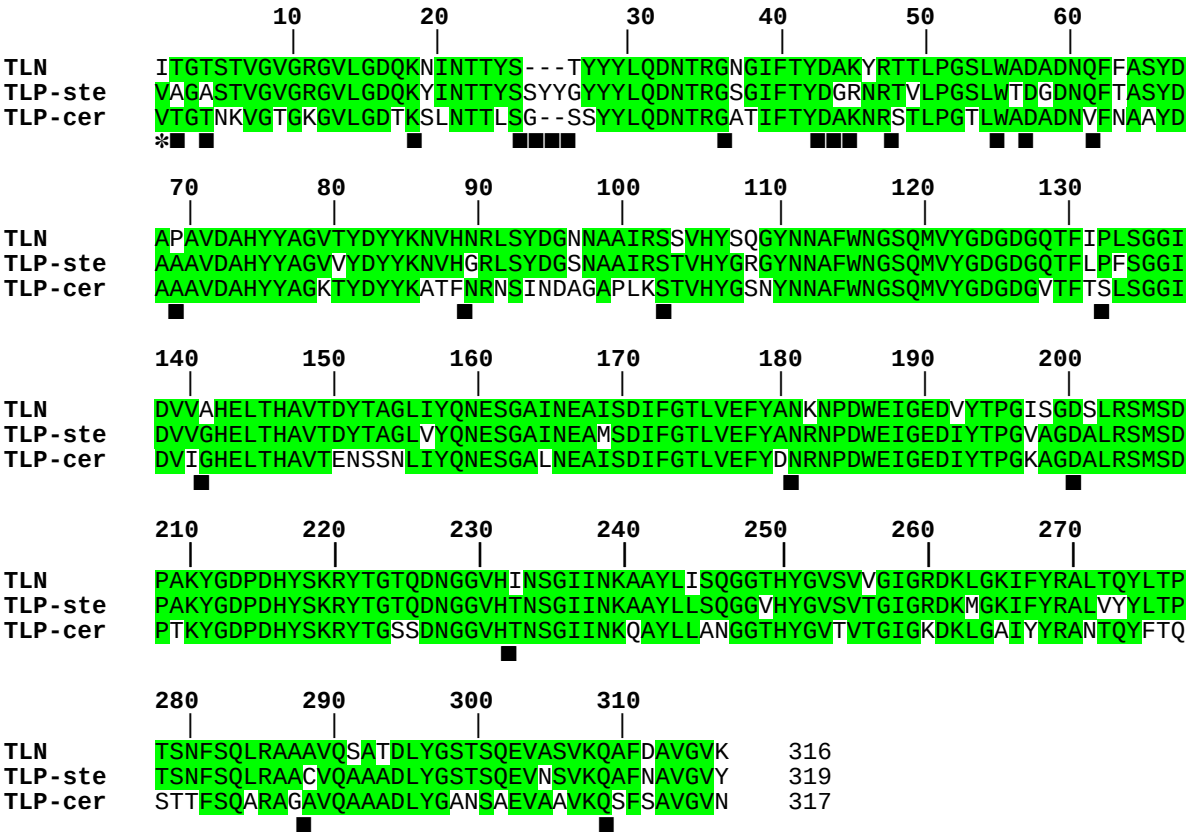


Figure 1. Alignment of the amino acid sequences of thermolysin (TLN, O'Donohue *et al*., 1994) and the TLPs produced by *B. stearothermophilus* (TLP-ste; Takagi *et al.*, 1985) and *B. cereus* (TLP-cer; Wetmore *et al*., 1992). The residues are numbered according to the sequence of TLN. ■, Substitutions between TLN and TLP-ste that were analysed by mutagenesis. ✻, Substitutions that were not analysed by mutagenesis. , Substitutions that were not analysed by mutagenesis; other mutations at these positions have been described previously (see text for further details).

exclusively affected by surface-located mutations and that these mutations are clustered in certain parts of the protein (Eijsink *et al.*, 1995).

The thermostable TLN retains 50 percent of its activity during 30 minutes of incubation at 86.9°C (T_{50}=86.9°C). TLN differs at 43 positions from the less thermostable TLP produced by *B. stearothermophilus* CU21 (T_{50} = 73.4°C). An alignment of these two proteins is shown in Figure 1. A three-dimensional model of TLP-ste was built on the basis of the crystal structure of thermolysin (Holmes & Matthews, 1982; Vriend & Eijsink, 1993). Considering

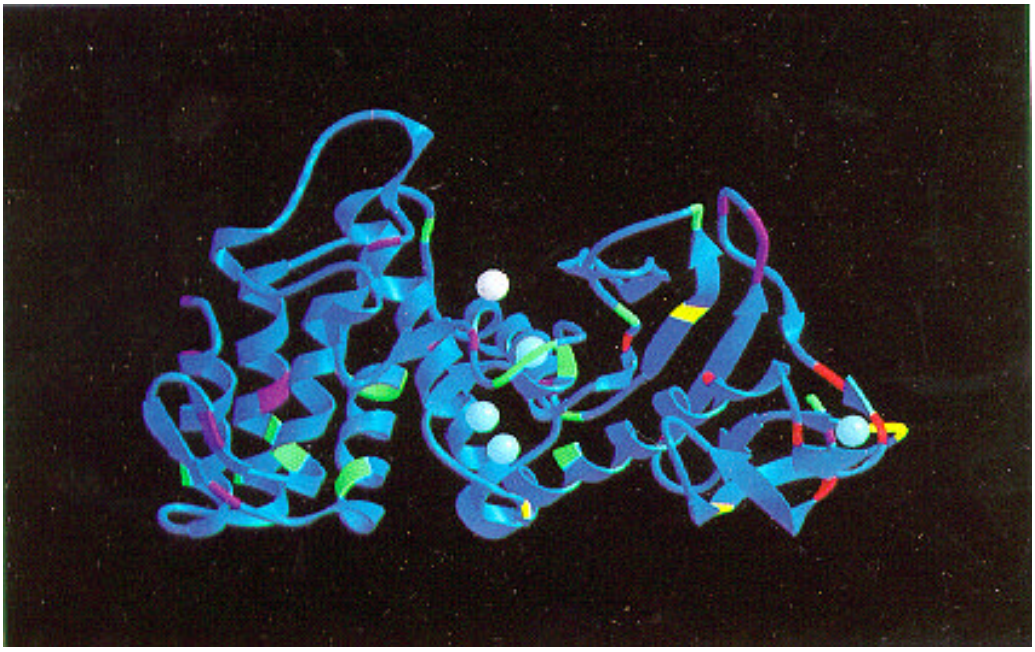


Figure 2. Ribbon drawing of TLP-ste. Positions at which TLP-ste differs from TLN (see Table 2 and 3) are shown in green (no mutation results available), purple (mutated; absolute effect on T_{50} less than 1 C), yellow (mutated; destabilising effect of at least 1 C), or red (mutated; stabilising effect of at least 1 C). The active site zinc (green sphere) is located in the active site cleft, between the -pleated N-terminal domain (right) and the largely α-helical C-terminal domain (left). Calcium ions are indicated by white spheres. The calcium bound in the N-terminal domain is calcium-3. The picture was obtained using the program Ribbons (Carson, 1991).

the high identity between thermolysin and TLP-ste (86%), the model of the latter is expected to be quite accurate (Sander & Schneider, 1991; Mosimann *et al.*, 1995). This is illustrated by the fact that the TLP-ste model has been used for the successful design of stabilising mutations, comprising for example the improvement of hydrogen bonding networks (Vriend & Eijsink, 1993) and the introduction of a drastically stabilising disulphide bridge (Mansfeld *et al.*, 1995). All residues shown to be involved in catalysis and the binding of the zinc and calcium ions (Holmes & Matthews, 1982; Hangauer *et al.*, 1984) are conserved between the two enzymes.

In previous studies, several 'TLP-ste → TLN' mutations were analysed. Some of these mutations had strong stabilising effects, whereas others only contributed marginally to thermal

stability of the enzyme (Eijsink *et al.*, 1995). Combining these stabilising mutations resulted in TLP-ste variants that were more stable than TLN itself, suggesting that some of the other 'TLP-ste → TLN' replacements are destabilising. In the present study we present a complete overview of the structural differences between TLP-ste and TLN. The importance of differences between the two enzymes with respect to stability was evaluated using molecular modelling and conclusions drawn from previous site-directed mutagenesis studies. Differences that seemed to be important on the basis of these analyses were evaluated experimentally by analysing the effects of the corresponding amino acid substitutions on the stability of TLP-ste.

Several of the important mutations described in this report are located near the binding site of one of the four calcium ions in TLP-ste and TLN (Calcium-3; Holmes & Matthews, 1982; Figure 2). To investigate whether these mutations exert their effect on stability via modulation of the binding of Calcium-3, the effect of the calcium concentration on T_{50} was also addressed in the present study.

MATERIALS AND METHODS

Production and characterisation of mutants

The gene encoding the *nprT* gene of *B. stearothermophilus* was originally cloned and sequenced by Aiba and co-workers (Fujii *et al.*, 1983, Takagi *et al*., 1985) and subcloned as described previously (Eijsink *et al.*, 1992c). Plasmid pCO3, a Bacillus vector containing the gene of *Bacillus cereus* encoding TLP-cer, was a gift of Dr. R. Roche, Calgary, Canada (Wetmore *et al.*, 1992). The gene encoding TLP-sub (Yang *et al.*, 1984) was a gift of Gist-Brocades N.V., Delft, The Netherlands, and subcloned as described elsewhere (Eijsink *et al*., 1992a). Cloning, sequencing and expression of the gene encoding TLP-cal have been described by Van den Burg *et al.* (1991).

All methods for side-directed mutagenesis of TLP encoding genes and for the expression of these genes in *B. subtilis* DB117 have been described elsewhere (Eijsink *et al.*, 1992c). Production, purification and subsequent characterisation of wild-type and mutant TLPs were performed as described earlier (Van den Burg *et al.*, 1989; Eijsink *et al*., 1992c). All wild-type enzymes listed in Table 1, except TLN, were obtained via over-expression of the corresponding genes in *B. subtilis* DB117 and subsequent purification of the gene product. Partially pure TLN was obtained from Boehringer Mannheim, Germany and further purified using the standard protocol.

For the determination of thermal stability 0.1 M solutions of purified protease (in 20 mM sodium acetate, pH 5.3, 5 mM $CaCl_2$, 0.01% triton X-100, 0.5% isopropanol, 62.5 mM NaCl) were incubated at various temperatures for 30 minutes, after which the residual proteolytic activity was determined using casein as a substrate (Fujii *et al.*, 1983). Thermal stability was quantified by the T_{50}, being the

temperature giving 50% residual protease activity. A sample of wild-type protease was included in each assay, and the stabilities of the mutants were expressed as T_{50}, being the difference in T_{50} between the mutant and the wild-type enzyme. T_{50} values given are the average of at least three independent assays with an error margin of approximately 0.3°C. The effect of calcium on the T_{50} was determined by varying the concentrations of $CaCl_2$ in the buffer described above.

Structural analysis

The originally published primary structure of TLN, based on amino acid sequencing, contained an Asp at position 37 and a Glu at position 119 (Titani *et al.*, 1972). Kubo & Imanaka (1988), Nishiya & Imanaka (1990) and O'Donohue *et al.* (1994) determined the DNA sequences of genes from *B. stearothermophilus* MK232, *B. stearothermophilus* TELNE and *B. thermoproteolyticus* rokko, respectively, that encode identical TLPs whose sequences differ from that of TLN by having Asn and Gln at positions 37 and 119, respectively. On the basis of these observations we assume in the present study that the crystallised and sequenced TLN in 1972 (Matthews *et al.*, 1972; Titani *et al.*, 1972) contains Asn and Gln at positions 37 and 119, respectively.

A three-dimensional model of TLP-ste was built on the basis of the crystal structure of TLN (Holmes & Matthews, 1982) using WHAT IF (Vriend, 1990), as described in detail previously (Vriend & Eijsink, 1993). Considering the high sequence identity (86 %) between the two enzymes, the model of TLP-ste is expected to be sufficiently reliable to predict and analyse the effects of site-directed mutations in general (Sander & Schneider, 1991; Vriend & Eijsink, 1993; Mosimann *et al.*, 1995). This was confirmed by the fact that the model has been used successfully for the *de novo* design of stabilising mutations (Imanaka *et al.*, 1986; Vriend & Eijsink, 1993; Mansfeld *et al.*, 1995). TLP-ste contains a stretch of three additional residues in the 25-30 region. This insertion (which in this study is regarded as a single difference between TLP-ste and TLN; see below) could not be modelled accurately and was omitted from the TLP-ste model. The approximate location of the insertion was however included in the evaluation of the structural effects of amino acid substitutions, as described below. In this report residues are numbered according to the sequence of TLN (Figure 1).

Several methods implemented in WHAT IF (Vriend, 1990) were used for comparison of TLP-ste and TLN and used for a detailed analysis of the structural effects of the differences between the two enzymes. One of the most important tools was the use of database-oriented techniques such as position-specific rotamer searches (Vriend & Eijsink, 1993; Vriend *et al.*, 1994). The use of position-specific rotamers was previously shown to be valuable for prediction of local structural changes that result from the introduction of a point mutation (De Filippis *et al.*, 1994). Water accessible surface areas were defined according to Lee & Richards (1971) and calculated using a WHAT IF module that gives approximately the same results as the Connolly surface-area-calculation package (Connolly, 1983). Side chain volumes were calculated according to Creighton (1993).

RESULTS AND DISCUSSION

Global description of the differences between thermolysin and TLP-ste

Table 1 shows the T_{50} values of five different wild-type TLPs, illustrating the differences in stability between naturally occurring TLPs. An overview of the sequence differences between TLP-ste and the two TLPs for which crystal structures are available is shown in Figure 1. The differences between TLP-ste and TLN are described in structural terms in Table 2. As illustrated in Figure 2, these differences between the two enzymes are scattered throughout the molecule. In accordance with observations made for other protein families (e.g. Fr mmel & Sander, 1989; Wilson *et al.*, 1990; Rentier-Delrue *et al.*, 1993; Serrano *et al.*, 1993), the large majority of the differences concern surface-located residues (Table 2, Figure 2). All residues known to be directly involved in binding of the zinc ion and catalysis (Holmes & Matthews, 1982) are conserved between TLN and TLP-ste. All residues directly involved in binding the four calcium ions are also conserved, but several differences occur in the direct environment of the conserved Asp57 and Asp59 which bind calcium-3.

The 43 'TLP-ste→ TLN' amino acid differences listed in Table 2 show a few general trends. In TLN there is an increase in side chain polarity and a small increase in side chain volume (by 104 A^3). Modelling studies indicated that the differences in the core of the proteins do not result in significant changes in the volume of internal cavities. Most of the latter substitutions are expected to deteriorate the internal hydrophobic packing (for example Val80→ Thr, Cys288→ Ala, and Ala291→ Ser; see Table 2 for details). Five mutations concern charged residues (Arg45→ Lys, Arg108→ Gln, Arg182→ Lys ; Asn311→ Asp; Tyr316→ Lys). Only the replacement at position 311 has a clear effect (increase) on the number of salt bridges (see Table 2).

Comparisons at the level of amino acid composition show only a few systematic differences. Going from TLP-ste to TLN, the number of glycines is reduced by six (from 42 to 36). Furthermore, there is an increase in the number of isoleucines (from 12 to 18) with a concomitant decrease in the number of valines (from 27 to 22). In TLN, the number of arginines is reduced by three (from 13 to 10), whereas the number of lysines is increased by

Table 2. Structural analysis of the 43 differences between TLP-ste and thermolysin.

	TLP-ste	TLN	Cß-$H_2$0 ( )[1]	Access. surface (^{2}) [2]	SS[3]	Interactions
1	Val	Ile	4.0	31	E	Exposed side chain; effect on stability unclear; possibly hydrophobic contacts between Ile-C and phenyl ring of Tyr29 in TLN only; possibly interaction with 25-30 region[4] in TLP-ste.
2*	Ala	Thr	3.3	37	S	Exposed side chain without contacts at the very beginning of a ß-strand; ß-branched Thr gives main-chain rigidification.
4*	Ala	Thr	3.2	21	S	Exposed side chain in ß-strand, ß-branched Thr gives main chain rigidification.
19*	Tyr	Asn	3.6	29	S	Exposed side chain covering the side chains of Val7 and (in the case of Tyr) Val9 contacts; TLP-ste configuration seems favourable; see figure 7.
25-30*			[4]	[4]	[4]	Three residue insertion in TLP-ste; see text and figure 8.
37*	Ser	Asn	3.5	26	S	Exposed side chain without contacts; structural differences between TLN structure and TLP-cer structure; effects on stability unclear.
44*	Gly	Ala	6.9	0.8	E	Possible rigidification by Gly→ Ala; Ala-C buried in TLN; contacts with non-conserved residue 103; see text and figure 8.
45*	Arg	Lys	3.3	35	E	Exposed side chain; both Arg and Lys could make various hydrogen bonds; local structural differences between TLN and TLP-cer (also Lys at 45); effects on stability unclear.
46*	Asn	Tyr	4.1	31	S	Exposed side chain with several possibilities for weak hydrogen bonds; partly dependent on residue 45; effects on stability unclear.
49*	Val	Thr	3.6	29	S	Exposed hydrophobic side chain in TLP-ste with a few hydrophobic contacts with the aliphatic

						part of Arg47; Thr in TLN is less hydrophobic but side chain has no contacts; counteracting effects on stability.
56*	Thr	Ala	4.0	13	S	Ala-Cß has no contacts in TLN; Thr in TLP-ste may have contacts with the 25-30 insertion[4]; Thr may reduce main-chain flexibility in TLP-ste; effects on stability unclear, see text.
58*	Gly	Ala	3.3	22	S	Ala-Cß is exposed and has no contacts; main-chain rigidification by Gly→ Ala[5].
63*	Thr	Phe	3.3	23	S	Phe results in hydrophobic clustering at the surface; studied in detail previously[6]; see text.
69*	Ala	Pro	5.5	1.6	H	In the N-terminal turn of an α-helix; almost completely buried side chain; - angles are favourable for Pro[7,8]; TLP-cer (Ala at 69) has same structure as TLN; main effect of Ala→ Pro is rigidification.
80	Val	Thr	8.7	0	H	Buried side chain; Thr-O 1 makes only one weak hydrogen bond in TLN; Val gives better hydrophobic packing.
89*	Gly	Asn	3.2	38	S	Exposed side chain; Asn side chain without clear interactions in TLN; rigidification by Gly→ Asn.
96	Ser	Asn	4.3	35	E	Exposed side chain with the potential to form weak hydrogen bonds in both TLN and TLP-ste.
103*	Thr	Ser	8.6	0	S	Buried side chain; Thr in TLP-cer; Ser-O and Thr-O 1 make similar hydrogen bonds (with N 1 of conserved Gln119 and bb-O of conserved Pro51); C 2 of Thr fills a cavity large enough to accommodate a water molecule; correlated with the mutation at position 44; see text and figure 8.
107	Gly	Ser	3.9	6.7	S	Exposed Ser side chain has no contacts in TLN. TLP-cer (Gly at 107) locally has the same conformation as TLN; rigidification by Gly→ Ser.
108	Arg	Gln	3.3	44	S	Exposed side chain without contacts.
131	Leu	Ile	3.7	13	S	Exposed side chain; Leu and Ile have similar hydrophobic interactions with C 1 of Val13 and C 1 of Leu14 in the N-terminal -hairpin.
133*	Phe	Leu	9.8	2.1	E	Buried side chain; Phe gives better hydrophobic packing but is forced into an unfavourable rotamer; counteracting effects, see text.
141*	Gly	Ala	10.8	0	H	Buried residue in α-helix. Ala-C buried in the same cavity as Val80; rigidification by Gly→ Ala[5] seems to be the main effect.

156	Val	Ile	4.1	24	E	Exposed side chain without contacts.
168#	Met	Ile	12.8	0.2	H	Buried side chain; Met almost the same size as Ile; both side chains fit; Met and Ile have similar amounts of hydrophobic contact.
182*	Arg	Lys	3.5	13	S	Interdomain salt bridge with Asp16, which is stronger in the case of Arg; see text and figure 5.
192	Ile	Val	7.7	0.5	E	Buried side chain; Ile gives better hydrophobic packing than Val.
197	Val	Ile	3.9	33	S	Exposed side chain with no contacts.
198	Ala	Ser	3.2	29	S	Exposed side chain with no contacts.
201*	Ala	Ser	6.0	0.8	S	Ala-C resides in a hydrophilic cavity (figure 3); a Ser side chain would fill that cavity, simultaneously improving the local hydrogen bonding network; Ser seems clearly favourable (figure 3).
232*	Thr	Ile	5.1	13	E	Thr→ Ile results in increased exposure of hydrophobic surface; Ile has favourable hydrophobic contacts with conserved Tyr221; counteracting effects.
244	Leu	Ile	9.7	0.2	H	Buried side chain, which is part of the same hydrophobic cluster as (the non-conserved) 263; Leu244 - Met263 packs equally well as Ile244 - Leu263.
249#	Val	Thr	3.3	27	S	Exposed side chain; located in a 10 amino acid ß-hairpin[9]; in TLN Thr-O 1 forms a hydrogen bond with Ser254-O in the same -hairpin; see text.
256#	Thr	Val	4.3	26	S	Exposed side chain; located in the same ß-hairpin as 249[9]; in TLP-ste Thr-O 1 forms a hydrogen bond with Gln308-O 1.
263	Met	Leu	9.1	0.5	H	Buried side chain; both Met and Leu fit; see residue 244.
272	Val	Thr	3.3	11	E	Exposed side-chain with weak contacts mainly with 273. Val272-Tyr273 in TLP-ste seems approximately equal to Thr272-Gln273 in TLN.
273	Tyr	Gln	3.5	33	E	Exposed side chain without close contacts; see residue 272.
288*	Cys	Ala	9.0	0	H	Buried side chain; Cys→ Ala creates a hydrophobic cavity (volume) and reduces hydrophobic interactions.
291	Ala	Ser	8.2	0	H	Ser-O is buried and makes only one weak H-bond with 287-bb-O; Ser→ Ala is favourable in α-helix.

293	Ala	Thr	4.4	12	H	Thr-O 1 does not make hydrogen bonds in TLN; Ala in α-helix is favourable compared to - branched Thr.
304	Asn	Ala	3.2	11	H	Asn has no interactions; both Asn and Ala are suitable residues in an α-helix.
311*	Asn	Asp	4.2	19	H	Asn with no contacts in TLP-ste; Asp in TLN forms salt bridge with (the conserved) Arg285; Arg285 side chain unsatisfied in TLP-ste. Asn→ Asp seems to be clearly stabilising.
316	Tyr	Lys	3.3	42	E	Exposed side chain with no contacts.

Substitutions were evaluated by comparing the crystal structure of TLN with a model of TLP-ste. Where needed and possible, the crystal structure of TLP-cer was used for additional analysis. *, Substitutions analysed by site-directed mutagenesis. #, Substitutions were not made, but mutations at these positions have been described previously; see text.

[1] The distance between a residue and the protein surface in the thermolysin three-dimensional structure is illustrated by the column Cß - H_2O, giving the shortest distance to a bulk water molecule of the ß-carbon atom in thermolysin (Vriend, 1990).

[2] Water accessible molecular surface area of the residue in thermolysin (Lee & Richards, 1971; Connolly, 1983).

[3] Secondary structure assigned with DSSP (Sander & Schneider, 1991); S, ß-strand; E, extended; T, turn; H, α- helix.

[4] TLP-ste: 25-SSYYGYYYL-30; TLN: 25-STYYYL-30; thus three residues are deleted and a Ser is changed into a Thr.

[5] Imanaka *et al.*, 1986.

[6] Van den Burg *et al.*, 1994.

[7] Hardy *et al.*, 1993.

[8] Watanabe *et al.*, 1994.

[9] Residues 249 and 256 are located in a -hairpin that can be deleted from TLP-ste without affecting stability (Eijsink *et al.*, 1992d).

three (from eight to 11).

Mutational analysis

All differences between TLP-ste and TLN were inspected and their importance was evaluated on the basis of general concepts with respect to protein stability. On the basis of structural inspections (Table 2) eleven out of the 43 TLP-ste→ TLN mutations were predicted to be stabilising (Tables 2,3). Six of these concerned Gly→ Xxx or Xxx→ Pro mutations, which, by their rigidifying effect on the unfolded protein, are expected to reduce local unfolding and the flexibility of the (partially) unfolded molecule (Matthews *et al.*, 1987). All presumably stabilising substitutions except Gly107→ Ser and Val249→ Thr were selected for analysis by site-directed mutagenesis. The Gly107→ Ser is located in a short surface loop and was omitted from the analysis because previous studies had shown that mutations in this surface loop had only marginal effects on stability (V.G.H. Eijsink, G. Vriend and G. Venema, unpublished results). The Val249→ Thr mutation was omitted because it is located in a ß-hairpin, the removal of which from TLP-ste was shown to have no effect on stability (Eijsink *et al.*, 1992d). The Gly44→ Ala mutation was tested in combination with two consecutive mutations at positions 45 and 46 (see below). As documented in Table 3, four of the presumably stabilising mutations indeed increased the stability of the TLP-ste, whereas the other seven mutations had only marginal effects.

Eight TLP-ste→ TLN mutations were predicted to be destabilising (Tables 2,3). Two of these (Tyr19→ Asn and Arg182→ Lys) were selected for mutagenesis studies because structural inspections indicated large destabilising structural effects and because they were expected to affect the stability of the important N-terminal domain of the TLPs. These two mutations indeed destabilised TLP-ste (Table 3). The Cys288→ Ala mutation in the hydrophobic core in the C-terminal domain had already been constructed for another purpose (Vriend & Eijsink, 1993) and is shown for completion. The Thr256→ Val mutation is located in a ß-hairpin the removal of which from TLP-ste was shown to have no effect on stability (Eijsink *et al.*, 1992d).

Thirteen of the TLP-ste→ TLN mutations were predicted to have no effects on stability, since no significant structural effects could be observed in modelling studies (Tables 2,3). Previous studies had already indicated that mutations at positions 168 and 316 had only marginal effects on the stability of TLP-ste (Eijsink *et al.*, 1992c,d; Vriend & Eijsink, 1993).

Table 3. Predicted and measured effects of replacing residues in TLP-ste by the corresponding residue of TLN[1].

Mutation	T_{50}[2] observed	Mutation	T_{50}[2] observed	Mutation	T_{50}[2] observed
Stabilisation predicted, Gly→Xxx and Xxx→Pro type:		Neutrality predicted,		Destabilisation predicted	
Gly44→Ala [3]		Ser96→Asn	n.d.	Tyr19→Asn	- 3.0
Gly58→Ala	+ 3.9	Arg108→Gln	n.d.	Val80→Thr	n.d.
Ala69→Pro	+ 6.3	Leu131→Ile	n.d.	Arg182→Lys	- 2.9
Gly89→Asn	- 0.3	Val156→Ile	n.d.	Val192→Ile	n.d.
Gly107→Ser	n.d.	Met168→Ile	≈ 0[4]	Thr256→Val	≈ 0[4]
Gly141→Ala	+ 0.7	Val197→Ile	n.d.	Cys288→Ala	0.0
		Ala198→Ser	n.d.	Ala291→Ser	n.d.
		Leu244→Ile	n.d.	Ala293→Thr	n.d.
		Met263→Leu	n.d.		
		Val272→Thr	n.d.		
Stabilisation predicted, various causes		Tyr273→Gln	n.d.	Unpredictable effects:	
Ala2→Thr	- 0.2	Asn304→Ala	n.d.	Val1→Ile	n.d.
Ala4→Thr	+ 1.9	Tyr316→Lys	≈ 0[4]	del 25-30 [5]	- 6.3
Ala201→Ser	- 0.4			Ser37→Asn	+ 2.4
Val194→Thr	≈ 0[4]			Arg45→Lys [3]	
Asn311→Asp	- 0.3			Asn46→Tyr [3]	
				Val49→Thr	- 0.3
				Thr56→Ala	+ 1.9
				Thr63→Phe	+ 7.0
				Thr103→Ser	- 6.0
				Phe133→Leu	+ 1.5
				Thr232→Ile	- 0.6

[1] Mutant TLPs were similar to wild-type TLP-ste with respect to production and purification yields. Specific activities towards casein were largely similar, but changes up to 50 % were observed for Phe133→Leu and Gly141→Ala.

[2] The error margin in T_{50} values is 0.3 C.

[3] The mutations at positions 44, 45, and 46 were combined. The T_{50} of the triple mutant was - 0.2 C.

[4] The results from previous mutagenesis studies strongly indicate that mutations at this position have marginal effects on stability; see text for details.

[5] Replacement of 25-SSYYGYYYL-30 by 25-STYYYL-30; thus, three residues were deleted and a Ser was changed into a Thr.

For various reasons the structural inspections did not provide clues regarding the effects of the last eleven differences between TLP-ste and TLN on stability (see Tables 2 and 3 for details). All of these except the rather conserved Val1→ Ile mutation were selected for further study by site-directed mutagenesis of TLP-ste. As illustrated by the T_{50} values given in Table 3, several of these mutations had considerable effects on stability.

The structural description of the various substitutions in Table 2 shows that only few of these substitutions are correlated via local contacts. Of the possibly correlated substitutions, the cluster involving residues 44 and 103 was investigated further by combining mutations, since this cluster included a mutation with a large effect on stability (Thr103→ Ser) and because detailed structural information for this cluster was available (see below). Combining the almost neutral G44A-R45K-N46Y mutations (T_{50} = -0.2 C) with the labile T103S mutant of TLP-ste (T_{50} = -6.0 C) increased the stability of the enzyme by 6.5 °C, to T_{50} = + 0.5 °C, indicating that the substitutions at the positions 44 and 103 are structurally correlated (Table 4).

Discussion of stability effects

A201S, N311D. Both our modelling studies and the results of previous mutagenesis studies indicated that the most important differences between TLP-ste and TLN (in terms of stability) are located in the N-terminal domains of the enzymes. Indeed, with the exception of the Arg 182→ Lys mutation, all mutations introduced in the C-terminal domain had marginal effects on stability. There are two 'TLP-ste→ TLN' substitutions in the C-terminal domain (Ala201→ Ser and Asn311→ Asp) that, without reservations, were predicted to be stabilising. Figure 3 shows that in TLN the side chain of Ser201 resides in a (conserved) hydrophilic cavity and that its hydroxyl group contributes considerably to the hydrogen bonding network in this cavity. Figure 4 shows how Asp311 in thermolysin forms a salt bridge with (the conserved) Arg285, thus providing a stabilising interaction between two C-terminal helices (consisting of residues 281-296 and 301-312). Considering the high sequence identity between TLP-ste and TLN in the structural environment of residues 201 and 311, the modelling of the effects of the mutations at these positions is likely to be accurate. Thus, the most probable explanation for the marginality of the mutational effects is that unfolding of the C-terminal domain is not rate-determining for the thermal inactivation of TLP-ste (see below).

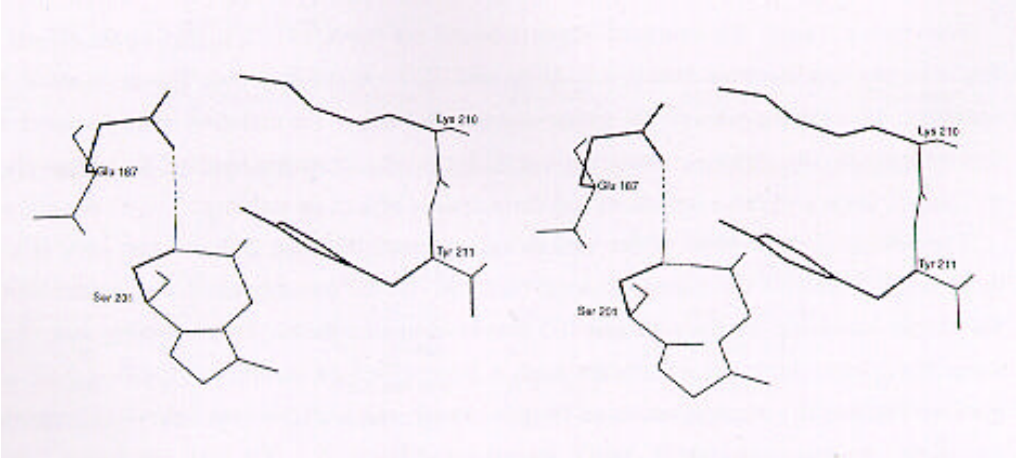


Figure 3. The environment of Ser201 in TLN. The side chain of residue 201 points into a hydrophilic cavity. The only hydrogen bond that can be made by the O 1 of Glu187 is the one with Ser201-O .

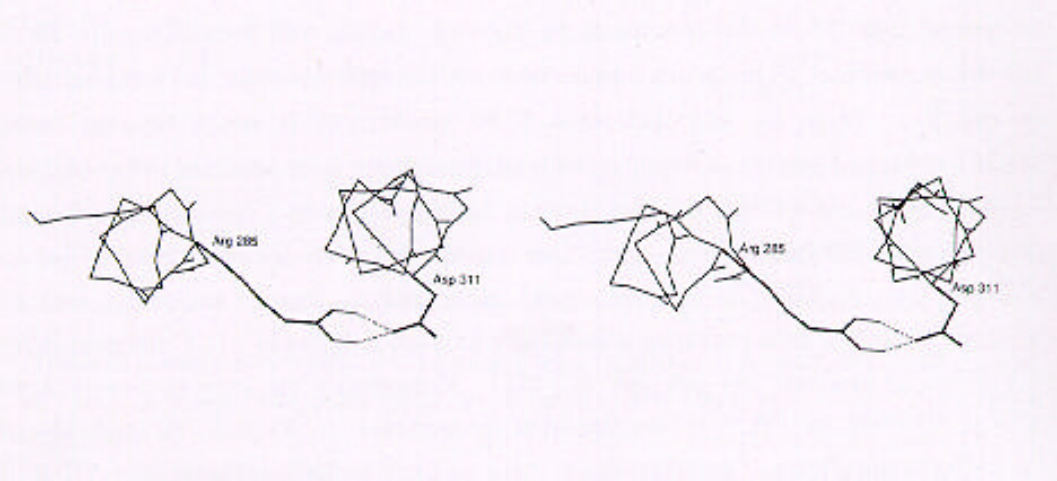
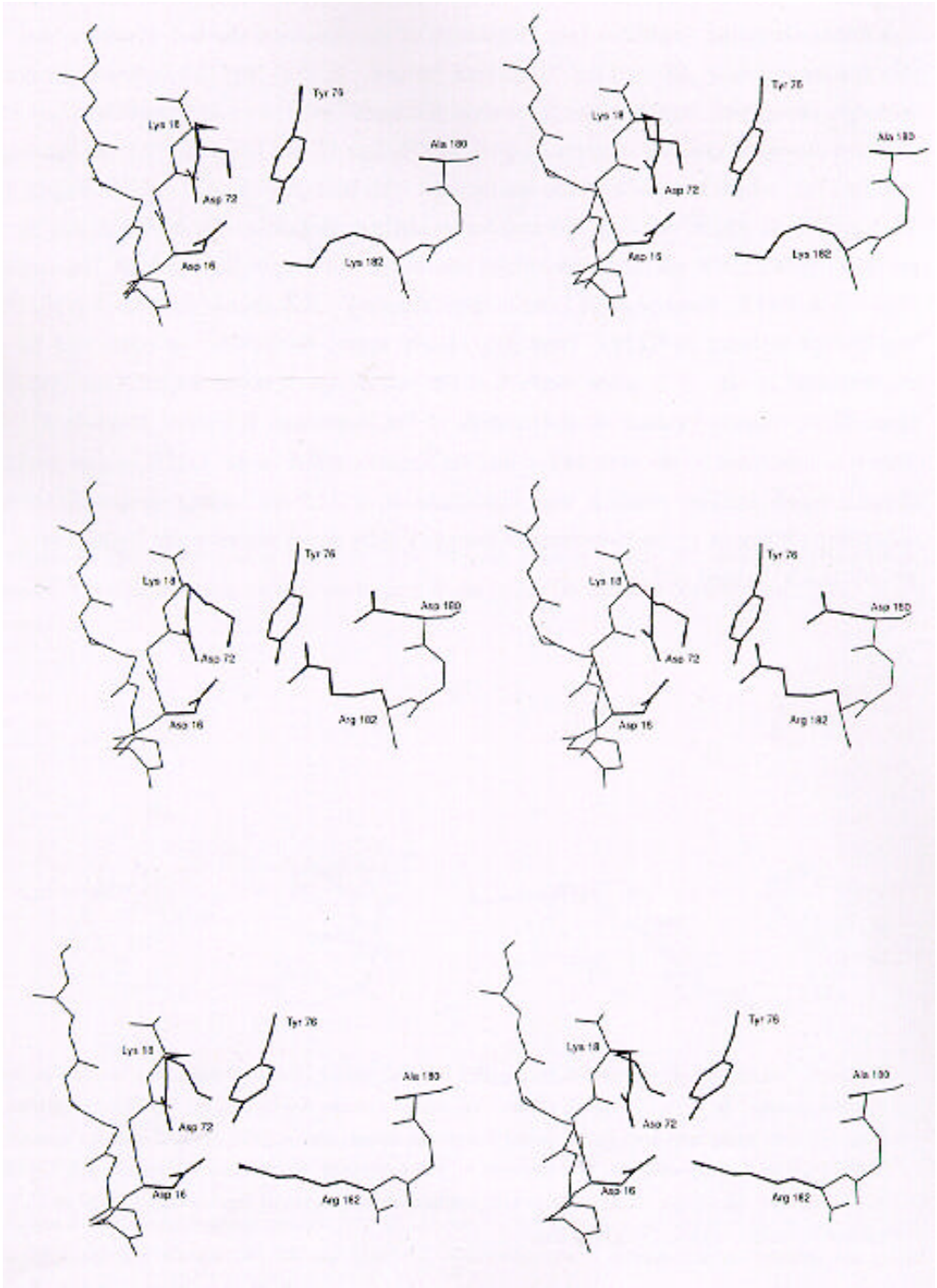


Figure 4. Salt bridge between residues Arg285 and Asp311 in TLN. This salt bridge connects two α-helices in the C-terminal domain (281-296 and 301-312). This interaction is weakened after replacing Asp by an uncharged Asn residue, as in TLP-ste.

R182K. The only mutation in the C-terminal domain with considerable effects on stability was the destabilising Arg182→ Lys. Inspection of the structures showed, however, that this substitution primarily affected the N-terminal domain. In TLN, Lys182 makes a salt bridge with (the conserved) Asp16 in the N-terminal ß-hairpin. As shown in Figure 5A, Asp 16 is part of a cluster of charged residues (Asp16, Asp72, Lys 18 and Lys182). TLP-cer has Arg at position 182, which has electrostatic interactions with both Asp180 and Asp16 (Figure 5B). TLP-ste has an Ala at position 180 and our modelling studies clearly indicated that in this enzyme Arg182 forms a strong interdomain salt bridge with Asp16 (Figure 5C). The Arg182-N 2 - Asp16-O 2 distance in TLP-ste is approximately 1.2 shorter than the Lys182-N - Asp16-O 2 distance in TLN. Thus Arg clearly seems favourable, as confirmed by the experimental result. It is quite likely that the salt bridge between Arg182 and Asp16 is favourable primarily because of stabilisation of the N-terminal ß-hairpin (residues 1 - 25). This was confirmed by the observation that the negative effect of the Arg182→ Lys mutation became much smaller when it was introduced in a TLP-ste variant containing several stabilising mutations in the N-terminal ß-hairpin and its direct environment (reduction of T_{50} by only 1.1 C instead of 2.9 C).

Figure 5. Interaction of residue 182 (C-terminal domain) with a cluster of conserved residues in the N-terminal domain. In TLN, Lys182-N makes a salt bridge with the Asp16-O 2 (a). In TLP-cer residue 182 is an Arg interacting with both Asp16 in the N-terminal domain and Asp180 in the C-terminal domain (b). TLP-ste has an Ala at position 180 and therefore Arg182 is likely to interact more exclusively with Asp16 (c). Arg182 in TLP-ste (c) has closer contacts with residues in the N-terminal domain than Lys182 in TLN (a); see text for further details. See next page.


Lys 18
Asp 72
Asp 16
Tyr 76
Ala 180
Lys 182
Lys 18
Asp 72
Asp 16
Tyr 76
Ala 180
Lys 182
Lys 18
Asp 72
Asp 16
Tyr 76
Asp 180
Arg 182
Lys 18
Asp 72
Asp 16
Tyr 76
Asp 180
Arg 182
Lys 18
Asp 72
Asp 16
Tyr 76
Ala 180
Arg 182
Lys 18
Asp 72
Asp 16
Tyr 76
Ala 180
Arg 182

A2T, A4T, T56A, T63F. The stabilising mutations Ala4→ Thr, Thr56→ Ala and Thr63→ Phe were originally suggested by a comparison of the TLP from *B. caldolyticus* (TLP-cal) with TLP-ste (Van den Burg *et al.*, 1991). The stabilising effect of the most prominent of these mutations, Thr63→ Phe, was rather unexpected, since it causes the replacement of a hydrophilic Thr by a hydrophobic Phe at the protein surface. The remarkable stabilising effect of this mutation has been studied in detail and it was shown that hydrophobic interactions between the side chain of residue 63 and the aliphatic parts of protruding side chains in its direct environment (Val9, Arg11, Gln17, Gln61; all conserved between TLP-ste and TLN) contributed to stability (Van den Burg *et al.*, 1994: Frigerio *et al.*, 1996). The structural inspections did not provide an explanation for the stabilising effect of the Thr56→ Ala mutation and actually indicated that a Thr at position 56 would be favourable for stability. An unknown contact between residue 56 and the three extra residues present in TLP-ste could account for the unexpected effect of the Thr56→ Ala mutation. The stabilising effect of the Ala4→ Thr mutation in the N-terminal -hairpin is likely to be caused by the fact that the ß-branched Thr is more favourable for -sheet formation than Ala (e.g. Fassman, 1989; Kim & Berg, 1993; Otzen & Fersht, 1995; Finkelstein, 1995). In contrast to residue 4, residue 2 is not yet part of the N-terminal -sheet, which may explain why the analogous Ala2→ Thr mutation had only marginal effects on stability.

S37N. The Ser37→ Asn mutation is a supposedly neutral replacement of one fully exposed polar side chain by another. However, the local structure around residue 37 was found to differ significantly between TLN (Asn at 37) and TLP-cer (Ala at 37), indicating that mutations at position 37 might be important. The clear stabilising effect of the S37N mutation confirms this importance, but the reasons for this stabilisation remain unclear.

G58A, A69P, G89N, G141A. The rigidifying mutations Gly58→ Ala and Ala69→ Pro stabilised TLP-ste (Imanaka *et al.*, 1986; Hardy *et al.*, 1993). Comparison of the structures of TLP-cer (Ala at position 69) with thermolysin (Pro at position 69), showed that the Ala69→ Pro mutation has no major structural effects (Figure 6). Thus, rigidification seems to be the main effect of the Ala69→ Pro mutation, possibly augmented by a small improvement of hydrophobic contacts between the partially buried Pro and surrounding residues (Figure 6). Studies of the effects of additional mutations of residue 69 have confirmed that rigidification is the cause for the stabilising effect of Ala69→ Pro (see chapter 3). The rigidifying Gly141→ Ala and Gly89→ Asn mutations had only marginal effects on stability (discussed below).

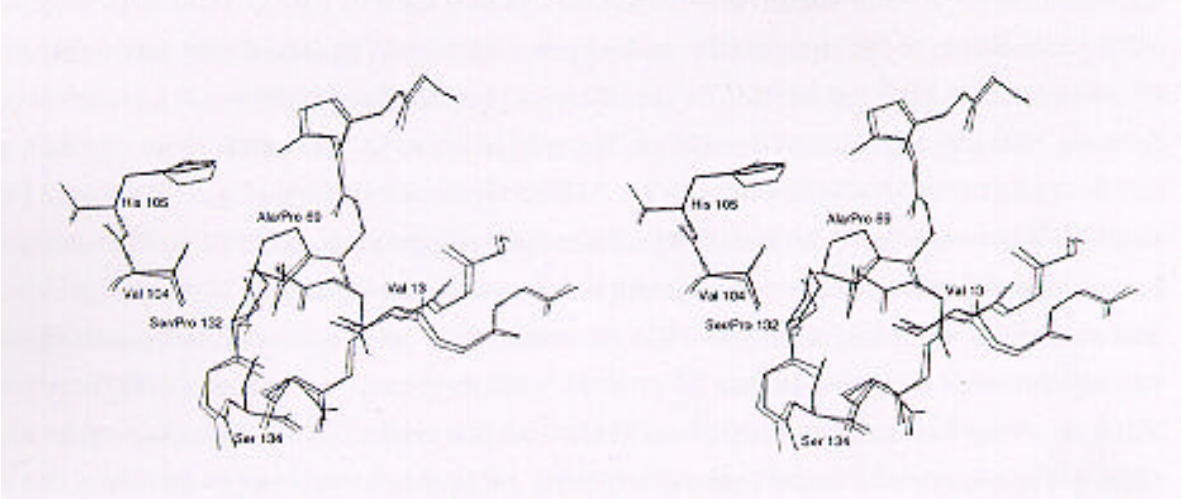


Figure 6. Superposition of the crystal structure of TLN (solid line) on the crystal structure of TLP-cer (dashed line) in the environment of residue 69. Residue 69 is positioned in the irregular N-terminal turn of the only α-helix in the N-terminal domain which extends to residue 87.

V49T, F133L, T232I. Structural analysis of the Val49→ Thr, Phe133→ Leu, and Thr232→ Ile mutations did not permit a prediction of the stability effect, since these mutations seemed to have both negative and positive effects (described in Table 2). Nevertheless, the observed effects on T_{50} indicate the relative importance of these opposing structural effects. In case of the stabilising Phe133→ Leu , the negative effect of forcing the Phe side chain into an unfavourable rotamer is apparently more important than the beneficial effect of increased packing by the Phe side chain.

N19Y. The side chain of residue 19 is solvent-exposed. Analogous contacts exist between Asn19-C and the C 1 of the exposed Val7 in TLN, and between Tyr19-C 1 (modelled in its most preferred rotamer) and Val7-C 1 in TLP-ste (Figure 7). Modelling studies indicated that

the situation in TLP-ste is more favourable for stability: there are better hydrophobic contacts between residues 19 and 7, and, furthermore, Tyr at position 19 also covers the hydrophobic side chain of Val9 (which is exposed in TLN). The hydrophobic interactions of the solvent-exposed Tyr19 resemble the favourable hydrophobic clustering around the surface located Phe63 (see above and Van den Burg *et al.*, 1994). In accordance with these considerations, the Tyr19→ Asn mutation destabilised TLP-ste.

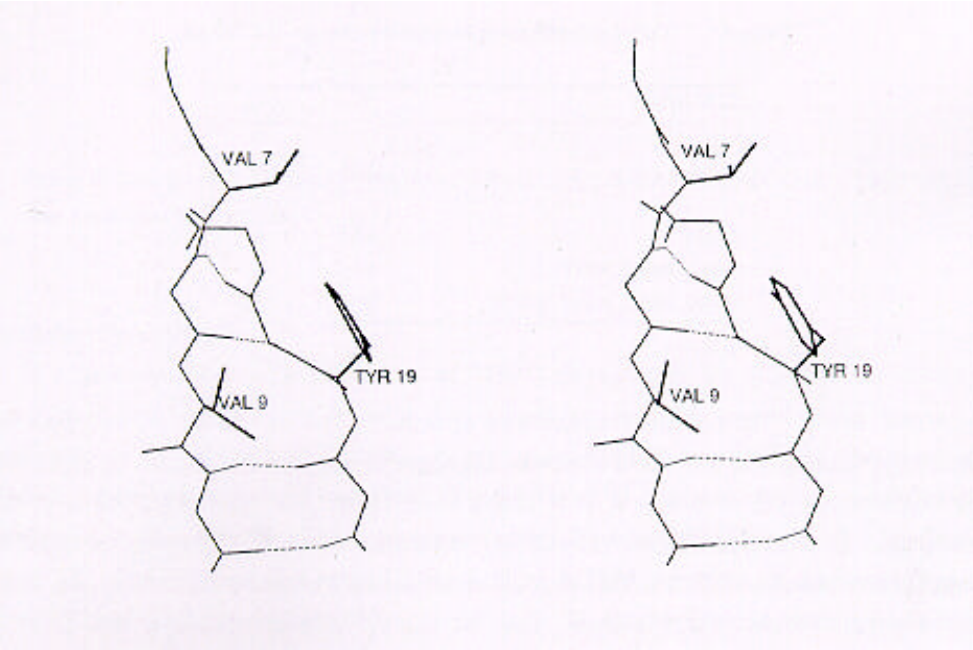


Figure 7. Interactions of the solvent exposed Tyr19 in the TLP-ste model. The residues 7, 9 and 19 are all located in the N-terminal -hairpin.Val7-C 1 has a hydrophobic contact with the C 1 of Tyr19. The side-chain of Val9, which is also part of the hydrophobic cluster around Phe63 (Van den Burg *et al.*, 1994), is protected from the solvent by the side chain of Tyr19.

Deletion between residues 25 and 30. The three amino acid insertion between residues 25 and 30 in TLP-ste could not be modelled and, thus, it is not easy to find a structural explanation for the large destabilising effect that was observed after deleting this insertion. It is conceivable that the two tyrosine residues in the insertion contribute to the stability of TLP-ste by covering exposed hydrophobic residues at the beginning of the N-terminal -hairpin, or in the environment of residue 56. Aromatic stabilisation (Burley & Petsko, 1985) could also play a role, since the insertion contributes to a remarkable clustering of tyrosine residues (Figure 1). It is conceivable that thermolysin can be stabilised by introducing the three extra residues present in TLP-ste.

Table 4. Correlated differences between thermolysin and TLP-ste.

TLP-variant	T_{50}	T_{50}
TLP-ste	73.4	0
T103S	67.4	- 6.0
G44A, R45K, N46Y	73.2	- 0.2
G44A, R45K, N46Y, T103S	73.9	+ 0.5

G44A, T103S. The combined replacement of residues 44 - 46 had a marginal effect on stability, presumably as a result of compensating negative and positive effects. As illustrated in Figure 8, the C of Ala44 in TLN fills a small cavity just below the surface of the molecule. In wild-type TLP-ste (Gly at 44) this cavity is filled by the C 2 of Thr103. Introduction of both Ala44 and Thr103 could result in steric hindrance. However, the TLP-cer structure (Ala44, Thr103) shows that the Ala-C and the Thr-C 2 can both be accommodated in the cavity, albeit with a rather small Ala-C - Thr-C 2 distance. Replacing Thr103 by Ser in TLP-ste (Gly at 44) creates a cavity, large enough to contain a water molecule, which may account for the large destabilisation that was obtained (T_{50} = - 6.0 C). In line with these considerations, the deleterious effect of the Thr103→ Ser mutation could be compensated by introducing the alanine at position 44, as illustrated in Table 4. In summary, these observations show that the mutations at positions 44, 45, 46 and 103 do not account for the stability difference between TLP-ste and TLN.

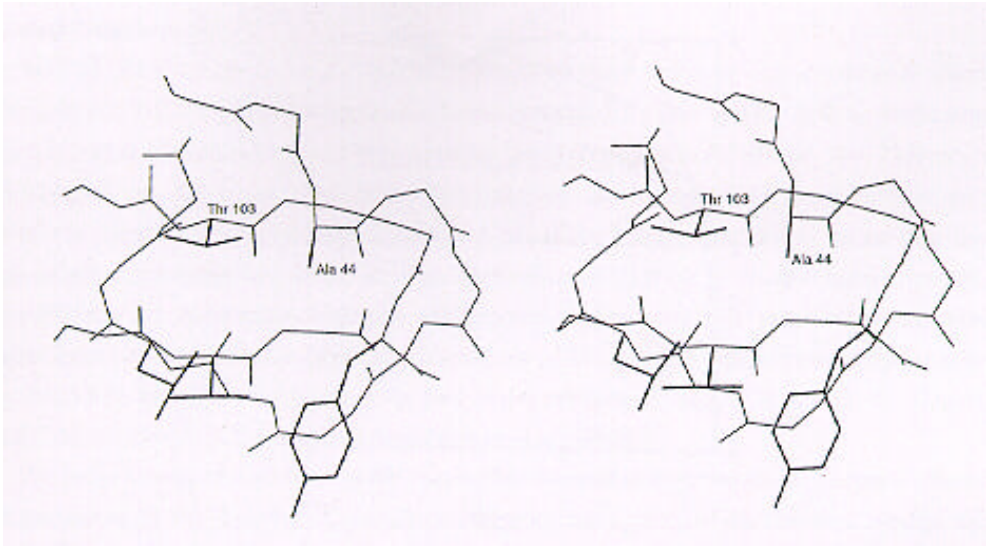


Figure 8. Environment of residues 44 and 103 in TLP-cer. The C-ß of Ala44 and the C- 2 of Thr103 fill the same cavity; see text for details.

Bridging the stability gap

In a previous publication, Eijsink *et al.* (1995) showed that the difference in stability between TLP-ste and TLN can be more than bridged by combining several of the stabilising mutations in the N-terminal domain. A mutant containing the combined mutations at positions 4, 56, 58, 63,and 69 (see Table 3) had a T_{50} of 93.3 C and addition of a designed mutation (Ser 65→ Pro) resulted in an enzyme with T_{50} = 96.9 C (TLP-ste: 73.4 C; TLN: 86.9 C). This suggests that some of the 'TLP-ste→ TLN' substitutions are destabilising, which is confirmed by the results of the present study. Similar observations were described by Serrano *et al.* (1993), who studied the structural determinants of a stability difference between two RNases by site-directed mutagenesis. Now that the gene encoding thermolysin has become available to the scientific community (O'Donohue *et al.*, 1994), it would be interesting to try to stabilise this enzyme by 'reverse' mutations such as Asn19→ Tyr, Lys182→ Arg, and the insertion between residues 25 and 30.

The stabilising Phe133→ Leu mutation was added to the sixfold mutant mentioned above,

which resulted in the most stable engineered TLP obtained so far, with a T_{50} of 98.1 C.

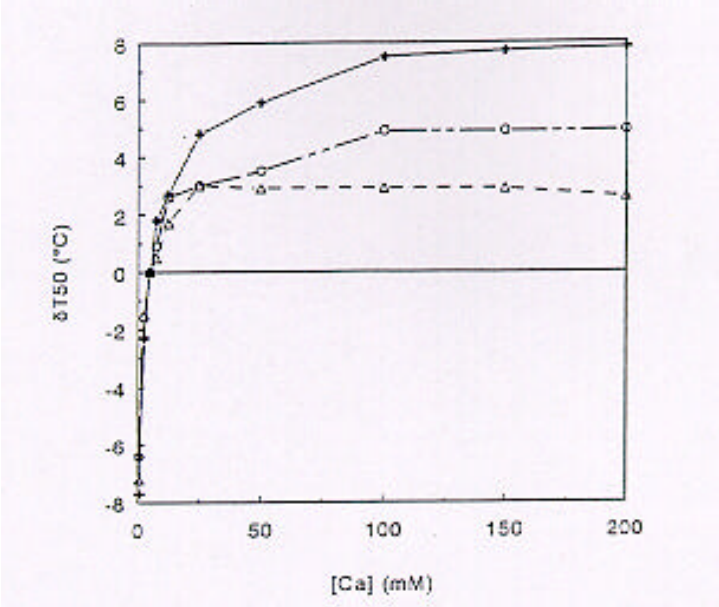


Figure 9. The effect of calcium concentration on thermal stability. The points indicate the T_{50} observed upon varying the $CaCl_2$ concentration. The T_{50} determined under standard conditions (5 mM $CaCl_2$) is taken as a reference point (T_{50} = 0). △, thermolysin (T_{50} = 86.9 C); +, TLP-ste (T_{50} = 73.4 C); ○, TLP-ste containing the mutations A4T, T56A, G58A, T63F, S65P, A69P (T_{50} = 96.9 C).

Effects of calcium

Since calcium ions are known to contribute to TLP stability (Matthews, *et al.*, 1972; Roche & Voordouw, 1978; Holmes & Matthews, 1982; Priest, 1989; Stark *et al.*, 1992) and since several mutations, with a high impact on stability, are close to the calcium binding site 3 (Figure 2), the effect of the calcium ion concentration on stability was studied. The results, presented in Figure 9, show that the stability of all variants tested depended on calcium concentration in a similar way. Apparently, the stability differences observed in wild-type TLPs and the effects of stability-determining substitutions are not accompanied by drastic changes in the contribution of the bound calcium ions. The results suggest that thermolysin has a slightly higher affinity for calcium than TLP-ste, since the former reaches maximum stability at lower $CaCl_2$ concentrations. This might be the result of differences in affinity for calcium-3, since the stable mutant (with mutations around the calcium-3 site) also reaches maximum stability at lower calcium concentrations. It is interesting to note that the sixfold mutant depicted in Figure 9 has a T_{50} of approximately 100 C in the presence of 50 mM $CaCl_2$.

Concluding remarks

Several of the mutations with significant effects on the stability of TLP-ste directly affect the N-terminal -hairpin made up by residues 1 through 25. This applies both to stabilising (4, 63) and to destabilising (19, 182, possibly the deletion between 25 and 30) 'TLP-ste→ TLN' mutations. Almost all other high-effect mutations are located in the direct environment of this N-terminal -hairpin. These observations confirm suggestions made earlier that the rate-limiting unfolding step in the thermal inactivation of TLP-ste is the N-terminal domain, in particular the N-terminal ß-hairpin. Supportive evidence for this proposed mechanism came from recent molecular dynamics simulations of thermolysin which showed that the N-terminal -hairpin is a relatively mobile part of the structure (D.M.F. Van Aalten, G. Vriend, H.J.C. Berendsen & V.G.H. Eijsink, unpublished observations).

The large effects of mutations in the N-terminal domain, as opposed to the marginal effects of mutations in the C-terminal domain corroborate the results of biochemical studies on domain stability of TLN. These studies showed that the C-terminal domain of TLPs has a high intrinsic stability and that unfolding of TLN is likely to start in the N-terminal domain (Corbett *et al.*, 1986; Vita *et al.*, 1989; Conejero-Lara *et al.*, 1994). This explains why mutations in the C-terminal domain that clearly seem to contribute to thermodynamic stability (which concerns global instead of partial unfolding) do in fact hardly affect the observed TLP stability (e.g. Ala201→ Ser, Asn311→ Asp; several other examples are described by Vriend & Eijsink, 1993). Likewise, the unexpected marginal effects of the Gly89→ Asn and the Gly141→ Ala mutation can only be explained by assuming that these residues are not involved in rate-limiting unfolding processes.

TLN and TLP-ste have high sequence identity (86%) and the differences between the two proteins are scattered over the molecule. Thus, every substitution resembles a point mutation, and can be expected to have only marginal effects on the protein structure (Matthews, 1993; De Filippis *et al.*, 1994). Although it cannot be excluded that the effect of mutations were misjudged in the modelling phase, table 3 shows that the predicted effects of individual 'TLP-ste→ TLN' mutations were reasonably correct. On the other hand, for some of the most important mutations the effects were not predictable and did not become apparent during extensive modelling studies, molecular dynamics simulations, or structural comparisons, although in some cases these studies *did* indicate that the effects could be considerable.

Several authors have analysed the structural causes of thermal stability by comparing the sequences and structures of homologous proteins differing in stability (e.g. Perutz & Raidt, 1975; Argos *et al.*, 1979; Men ndez-Arias & Argos, 1989; Fr mmel & Sander, 1989; Rentier-

Delrue *et al.*, 1993; V lkl *et al.*, 1994; Kornd rfer *et al.*, 1995; Russell & Taylor, 1995). These studies gave conflicting results and it is becoming more and more apparent that 'traffic rules' for stability can not be given, but that every single mutation should be evaluated in its specific context (Russell & Taylor, 1995). Some proposed 'traffic rules' have been used successfully in some cases (e.g. Mrabet *et al.*, 1992; Watanabe *et al.*, 1994). However, Serrano *et al.* (1993) compared two RNases by extensive site-directed mutagenesis studies and were unable to establish a clear correlation between the type of mutation and the effect on stability. In line with the observations by Serrano *et al.*, our results show that the difference in stability between TLP-ste and TLN depends on a small number of highly specific mutations that generally do not obey statistically derived 'traffic rules'. Thus our mutagenesis study confirms the suggestion made by Matthews *et al*. (Matthews *et al.*, 1974) that 'enhanced stability of thermostable proteins relative to thermolabile ones cannot be attributed to a common determinant such as metal ion or hydrophobic stabilisation, but in a given instance may be due to rather subtle differences in hydrophobic character, metal binding, hydrogen bonding, ionic interactions, or a combination of all of these'.

Acknowledgements:
We thank Bernard van der Vinne for technical assistance, Chris Sander for hospitality and stimulating discussions, and Ingolf F. Nes for general support.

References

References are listed in chapter 9