

## Prediction and analysis of structure, stability and unfolding of thermolysin-like proteases

Gert Vriend<sup>a,\*</sup> and Vincent Eijssink<sup>b,\*\*</sup>

<sup>a</sup>EMBL, Protein Design Group, Meyerhofstrasse 1, 6900 Heidelberg, Germany

<sup>b</sup>Department of Genetics, Center for Biological Sciences, Kerklaan 30, 9751 NN Haren, The Netherlands

Received 29 September 1992

Accepted 26 February 1993

*Key words:* Protease; Thermolysin; Stability; Unfolding; Model building

---

### SUMMARY

*Bacillus* neutral proteases (NPs) form a group of well-characterized homologous enzymes, that exhibit large differences in thermostability. The three-dimensional (3D) structures of several of these enzymes have been modelled on the basis of the crystal structures of the NPs of *B. thermoproteolyticus* (thermolysin) and *B. cereus*. Several new techniques have been developed to improve the model-building procedures. Also a 'model-building by mutagenesis' strategy was used, in which mutants were designed just to shed light on parts of the structures that were particularly hard to model.

The NP models have been used for the prediction of site-directed mutations aimed at improving the thermostability of the enzymes. Predictions were made using several novel computational techniques, such as position-specific rotamer searching, packing quality analysis and property-profile database searches. Many stabilizing mutations were predicted and produced: improvement of hydrogen bonding, exclusion of buried water molecules, capping helices, improvement of hydrophobic interactions and entropic stabilization have been applied successfully.

At elevated temperatures NPs are irreversibly inactivated as a result of autolysis. It has been shown that this denaturation process is independent of the protease activity and concentration and that the inactivation follows first-order kinetics. From this it has been conjectured that local unfolding of (surface) loops, which renders the protein susceptible to autolysis, is the rate-limiting step. Despite the particular nature of the thermal denaturation process, normal rules for protein stability can be applied to NPs. However, rather than stabilizing the whole protein against global unfolding, only a small region has to be protected against local unfolding. In contrast to proteins in general, mutational effects in proteases are not additive and their magnitude is strongly dependent on the location of the mutation. Mutations that alter the stability of the NP by a large amount are located in a relatively weak region (or more precisely, they affect a local unfolding pathway with a relatively low free energy of activation).

One weak region, that is supposedly important in the early steps of NP unfolding, has been determined in the NP of *B. stearothermophilus*. After eliminating this weakest link a drastic increase in thermostability was observed and the search for the second-weakest link, or the second-lowest energy local unfolding pathway is now in progress. Hopefully, this approach can be used to unravel the entire early phase of unfolding.

---

\*To whom correspondence should be addressed.

\*\*Present address: Laboratory of Microbial Gene Technology, P.O. Box 5051, 1432 Ås, Norway.

## INTRODUCTION

### *Model building by homology and design of site-directed mutations*

The availability of the 3D structure is a prerequisite for the rational design of site-directed mutations in a protein. X-ray crystallography and NMR spectroscopy are the only ways to obtain such detailed structural information. Unfortunately, these techniques involve elaborate technical procedures and many proteins fail to crystallize at all [1] and/or cannot be obtained or dissolved in large enough quantities for NMR measurements. The size of the protein is also a limiting factor for NMR. In the absence of measured data, model building on the basis of a known 3D structure of a homologous protein is at present the only reliable method to obtain structural information [2]. Comparisons of the tertiary structures of homologous proteins have shown that 3D structures have been better conserved in evolution than protein primary structures [3,4], indicating the feasibility of model building by homology.

Differences between 3D structures increase with differences in sequence identity and accordingly the accuracy of models built by homology decreases [2,5,6]. The errors in a model that is built on the basis of a protein with 90% sequence identity may be as low as the errors in crystallographically determined structures, except for a few individual side chains [5,7]. In case of 50% sequence identity RMS errors in the atom coordinates in the model can be as large as 1.5 Å, with considerably larger local errors. The deviations between a model built by homology and the 'real' X-ray structure vary through the molecule, the largest deviations occurring in surface-located loops. In well-defined secondary-structure elements and in the hydrophobic core, models are more accurate. The fact that the biggest modelling errors occur in solvent-exposed regions is of particular importance in the case of proteases, since it is now clear that exposed regions are the major determinants of the stability of this type of enzymes (e.g. [8–11]).

In recent years automatic model building by homology has become a routine technique which is implemented in most molecular-modelling software packages. Three main lines of research in this area are seen: (1) Improvement of the sequence alignment that is at the basis of the model building procedure; (2) Improvement of the way side chains are positioned in the modelled structure; and (3) Improvement of the way loops are inserted.

Several authors have described new techniques to align two proteins for cases where the structure of one of the proteins is known [12–22]. Rather than using the sequence of the known structure as input to the alignment procedure, a structure-dependent sequence profile is set up, and the sequence of the unknown structure is aligned with this profile. These methods will result in alignments that reveal less of the evolutionary path at the gene level, but are more suited for model building by homology. Also, these methods are potentially more powerful when aligning sequences in the 'twilight' zone (< 25% sequence identity).

Until recently side chains were normally placed using standard geometries [23,24], and a variety of procedures, ranging from manual adaptation [25] to molecular dynamics [26], have been used to get rid of van der Waals clashes. The latest developments are position-specific rotamers (see below) and Monte Carlo procedures combined with energy calculations [27].

The major problem in model building is still the insertion of loops. The main techniques used to model loops are:

- (1) Searching databases for loops with endpoints that match the points between which the loop has to be inserted in the structure to be modelled. This technique is based on the distance

geometry loop-searching algorithm originally described by Jones and Thirup [28]. Many programs use this technique (e.g., WHAT IF, QUANTA, Hydra, FRODO, Insight, O, BRAGI [29–34]).

(2) Ab initio building. This can be done in many ways, but a few points are common to all ab initio techniques: the fact that endpoints have to match is a very strong constraint. Further, van der Waals clashes have to be prevented, and all rules about proper stereochemistry and energetics should be obeyed. Many articles on these topics have been published, for example [35–41].

All modelling techniques depend at one stage or another on peptide or protein structure data. Our understanding of folding motifs etc. mainly comes from visual inspection of protein structures. The work of for example Chothia, Lesk and Richardson [42–45] has laid the foundations for this part of our knowledge. Most of what we know about protein geometry and energetics comes from studying small peptides (e.g. [46]).

Relational databases of protein 3D structures are now being used to improve our understanding of protein structures, and thus our possibilities in the field of model building. BIPED [47,48] and SESAM [49] are the first examples of operational relational protein sequence and structure databases. They are both based on commercial, general, relational database management systems. The latest database development is the property profile database [50]. This relational database management system was written especially to handle protein structure and sequence information. Its main advantage is that information can be extracted about residues in their environment in the protein, rather than about single residues. One of the more important advantages of the property profile database technique is that 3D information, like matching local structure fragments as described by Jones and Thirup [28], can easily be combined with sequence information to obtain position-specific rotamers. This method has been described by Levitt [51], and is implemented in a fully relational manner in WHAT IF [31,52]. Another important aspect is that directional atomic-contact analysis as described by Singh and Thornton [53] is available in a relational fashion. Directional contact analysis has been used to create a tool to evaluate the quality of packing of proteins [54]. When this packing analysis is combined with position-specific rotamers, mutants can be predicted automatically with high reliability (see below).

In protein engineering model-building techniques are used for the prediction of the structure of mutant proteins and, thus, for the prediction of the effects of designed site-directed mutations. Assuming the existence of a reliable 3D structure for the wild-type protein, and taking into account that site-directed mutations in most cases have very limited effects on the backbone structure of a protein [55,56], the modelling problem now is reduced to the prediction of one or a few side-chain conformations. This, however, does not make the problem less complicated, since for the proper evaluation of the energetics of a mutation — that is, for the evaluation of local side-chain interactions such as hydrogen bonds or hydrophobic contacts — highly accurate structural information is required [57]. Thus, for the rational design and analysis of site-directed mutations sophisticated modelling methods that are based on the usage of databases are essential. Molecular-dynamics (MD) simulations can be of help for a detailed structural analysis of a mutation, but the application of MD for the evaluation of the energetics of a mutation is highly complicated and of limited use [58].

#### *Bacillus neutral proteases*

Several members of the bacterial genus *Bacillus* produce extracellular metallo-endopeptidases,

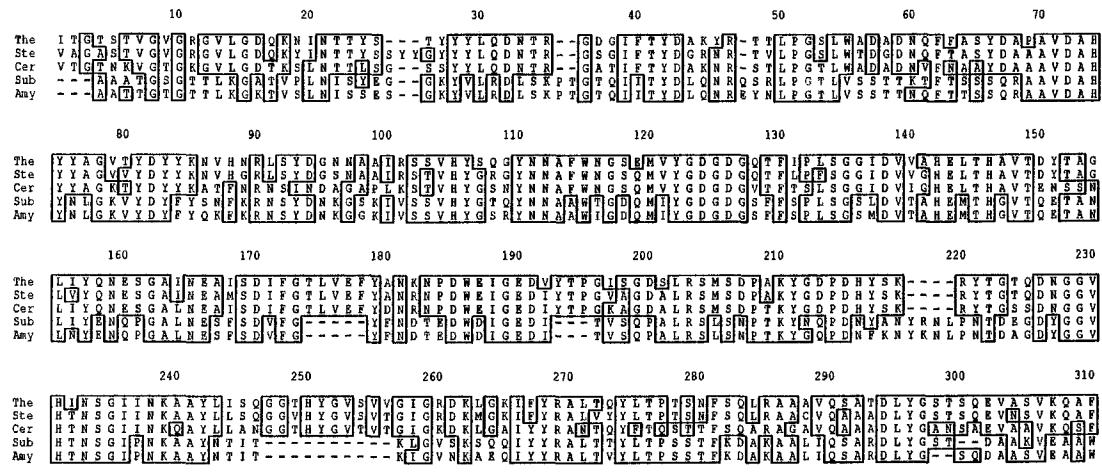


Fig. 1. Aligned amino acid sequences of the neutral proteases of *B. thermoproteolyticus* (The), *B. stearothermophilus* (Ste), *B. cereus* (Cer), *B. subtilis* (Sub), and *B. amyloliquefaciens* (Amy) [64–68]. The numbering of the residues refers to the sequence of thermolysin (The) and is used for all NPs throughout this paper.

called neutral proteases (NPs; E.C. 3.4.24.4). The NPs form a group of well-characterized homologous enzymes with similar enzymatic characteristics, such as binding of zinc and calcium ions, inhibition by chelating agents, similar pH optima, and a molecular mass of 32–35 kDa [59]. The NPs are initially synthesized as large precursor molecules, called pre-pro-proteins [60]. In the process of secretion, the pre- and pro-parts are cleaved off, and the mature enzymes are released into the culture medium [61]. NPs have a broad specificity [62,63] and the sites of cleavage in a globular protein substrate seem to be largely dictated by conformational features of the substrate molecule and not by sequence characteristics [11].

The amino acid sequences of several NPs are known (Fig. 1, Table 1 [8,64–69]) and for two of

TABLE 1  
HOMOLOGIES AND THERMOSTABILITIES OF NEUTRAL PROTEASES PRODUCED BY *BACILLUS* SPECIES<sup>a</sup>

Species	Name	Sequence identity	T <sub>50</sub> (°C)
<i>B. thermoproteolyticus</i>	thermolysin	100	82.0
<i>B. caldolyticus</i>	NP-cal	86	76.7
<i>B. stearothermophilus</i>	NP-ste	85	68.5
<i>B. cereus</i>	NP-cer	73	n.d.
<i>B. subtilis</i>	NP-sub	47	55.3
<i>B. amyloliquefaciens</i>	NP-am	47	n.d.

<sup>a</sup> Percentage sequence identities were calculated relative to thermolysin. T<sub>50</sub> values, indicating thermostability, were determined as described in the legend of Table 3. n.d. = not determined.

the enzymes the refined 3D structure, as determined by X-ray crystallography, is available (thermolysin, the NP of *B. thermoproteolyticus* [70,71] and NP-cer, the NP of *B. cereus* [72,73], see Fig. 2). The zinc ion is coordinated by the fully conserved residues His<sup>142</sup>, His<sup>146</sup> and Glu<sup>166</sup>. It plays a role in the catalytic mechanism, together with the conserved residues Glu<sup>143</sup>, His<sup>231</sup> and to a lesser extent Tyr<sup>157</sup> [74–79]. The NPs bind two (thermolabile variants) to four (thermostable variants) calcium ions, that contribute of the thermostability [80,81]. The highly thermostable thermolysin is by far the best characterized NP, and much is known about its structure, stability, activity, specificity and binding to inhibitors [11,79,81–83].

NPs are generally considered to consist of two domains (Fig. 2 [73,84]). The N-terminal domain consists mainly of  $\beta$ -structure. The C-terminal domain, comprising residues 155–316, can be divided into a middle (residues 155–241) and a C-terminal (residues 242–316) subdomain, which are connected by the  $\alpha$ -helix comprising residues 235–247. The middle subdomain has a rather irregular structure, dominated by a long irregular loop. The C-terminal subdomain is almost completely  $\alpha$ -helical and it has been shown to be stable by itself after cleavage from the enzyme by subtilisin or chymotrypsin [85].

Because they exhibit large differences in thermostability (Table 1), the NPs have been used for protein-engineering studies aimed at elucidating structural factors that determine protein thermostability [9,10,86–96]. Since until very recently the genes encoding thermolysin and NP-cer were not available [60,69], these studies had to be performed on NPs the crystal structures of which were not known. Thus, these studies were based on models built by homology.

#### *Stability and thermal inactivation of proteases*

The stability of a protein can be characterized by the free-energy difference  $\delta G_u$  between its folded and unfolded state [97–99]. Several methods for the analysis of reversible protein (un)fold-ing and the subsequent determination of  $\delta G_u$  exist [100–103]. Thermally induced reversible unfolding can be characterized by the  $T_m$  value. This is the temperature at which 50 percent of the protein molecules is unfolded (and in equilibrium with folded protein). The highest energy state that must be traversed in a reversible unfolding process is referred to as the transition state. Studies on the unfolding of several wild-type and mutant proteins have indicated that the confor-mation of the transition state resembles that of the folded state [99,104–110]. During unfolding, in the trajectory from the native to the transition state, most of the secondary structure appears to be preserved. Near the transition state a limited number of interactions is weakened to an extent that local conformational changes can occur. In the transition state the hydrophobic core of the protein seems to be weakened, but major conformational changes are restricted to localized regions outside the core [104,105,108,110].

Recent studies on nonnative cysteine bridge formation at room temperature in D-galactose chemosensory receptor [111] have shown that mobility of secondary-structure elements at the surface of molecules can take place with much larger amplitudes than previously assumed even under conditions in which global unfolding does not occur.

Proteases, especially a-specific ones like neutral proteases and subtilisins (serine proteases produced by various *Bacilli* [59]), differ from other proteins in that they often show autolysis. At elevated temperatures most subtilisins and NPs are irreversibly inactivated as a result of autolysis [112–115]. Preventing autolysis during stability measurements is highly complicated, if possible at all, and a rigorous thermodynamic analysis of protease stability therefore seems to be impossible

[112,116,117]. The use of inhibitors to prevent autolysis during unfolding studies is of limited use in this respect: first, it is complicated if not impossible to obtain a sufficient degree of inhibition, and second, it is hard to discriminate between the mass action effect of the inhibitor and the inhibition itself. The phenomenon of autolysis has substantial implications for analysing, measuring and engineering the thermostability of this class of enzymes.

Instead of using equilibrium measurements, the thermostability of proteases is usually quantified by kinetic analyses of the inactivation process. For example, instead of a thermodynamic  $T_m$ , a kinetic  $T_{50}$  can be used, which is the temperature at which, after a fixed incubation time, 50 percent of the initial protease activity remains. It has been hypothesized that reversible local unfolding processes which render the protease susceptible towards the irreversible process of autolysis determine its rate of inactivation (Fig. 3 [9,115]). Several observations support this view:

(1) Generally, the susceptibility of proteins towards proteolysis is increased when they are partially or fully unfolded [118–120]. For neutral proteases and for subtilisin it has been shown that autolysis is promoted under (slightly) denaturing conditions that probably induce local unfolding [112,121–123].

(2) Proteins with increased thermostability according to measurements of equilibrium unfolding are less susceptible to proteolysis [124–126].

(3) In several cases the thermal inactivation process of proteases was shown to be independent of protease concentration and catalytic efficiency and to follow first-order reaction kinetics [115,117].

(4) The inactivation rate of *B. subtilis* NP (NP-sub) at its  $T_{50}$  could be almost nullified by adding very high concentrations of a specific inhibitor to the enzyme, indicating that unfolding processes which occur at  $T_{50}$  have a reversible character. Apparently, nonautolytic irreversible inactivation processes, such as irreversible unfolding or deamidations, do not occur at this temperature.

(5) Considering what is known about the transition state of global unfolding (see above), local unfolding processes are supposedly taking place at the surface of the protein. Indeed, autolytic cleavage of thermolysin was found to occur predominantly in exposed regions of the protein, having high flexibility [11,127]. In addition, for the neutral protease of *B. stearothermophilus* (NP-ste) it was found that the most effective stabilizing site-directed mutations concerned surface-located parts of the molecule, and that even drastic changes in the core of the protein had only small effects on thermostability (see below [8,10]).

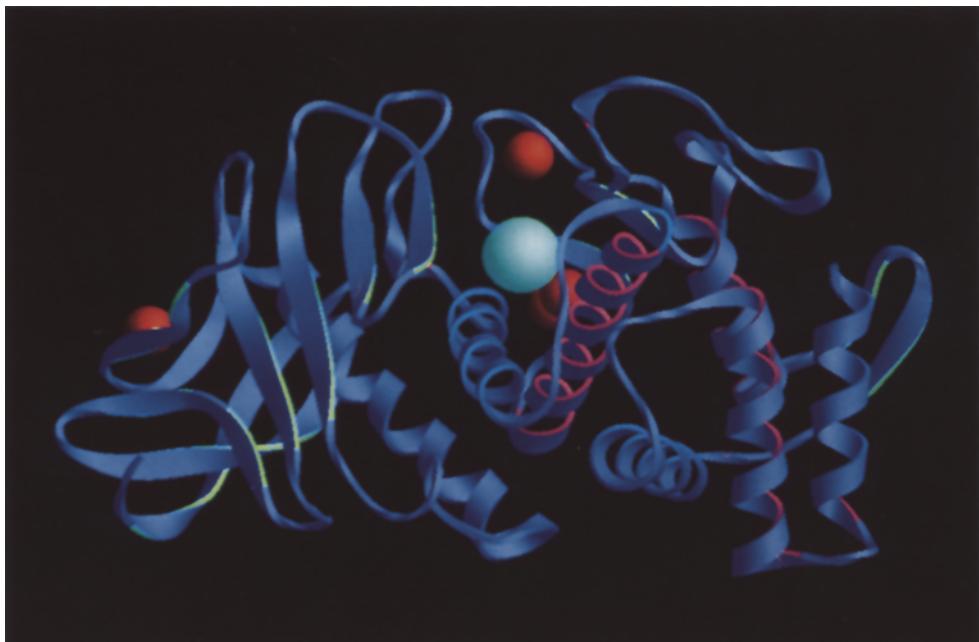
Some important consequences derived from the local unfolding model, such as nonadditivity of mutational effects and the existence of ‘sensitive’ regions at the protein surface have been confirmed by protein engineering experiments, which are discussed below.

## MODEL BUILDING OF *BACILLUS* NEUTRAL PROTEASES

### *Homology modelling of neutral proteases*

To obtain models of the NPs a fully automated procedure, implemented in the molecular-modelling program WHAT IF [31], was used. The refined structure of thermolysin [71] was taken as a starting point in which the appropriate residues were replaced. Where applicable, the crystal structure of NP-cer was used for the modelling of these residues. Side chains that had to be modelled because they occurred neither in thermolysin nor in NP-cer were appended to the

A



B

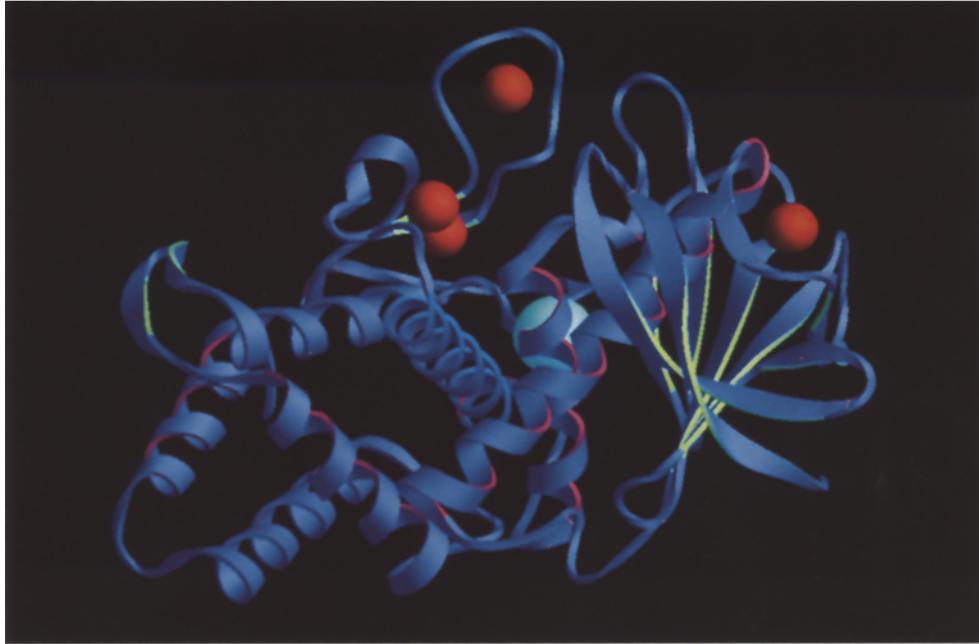


Fig. 2. Ribbon representations [215] of thermolysin, showing the zinc and calcium atoms as green and red spheres, respectively. The zinc atom is located in the interface between the N-terminal and C-terminal domain, near the active site. The N-terminal domain contains mainly  $\beta$ -sheets, whereas the C-terminal domain is largely  $\alpha$ -helical. (In A the N-terminal domain is on the left. B shows a rear view and thus the N-terminal domain is on the right).

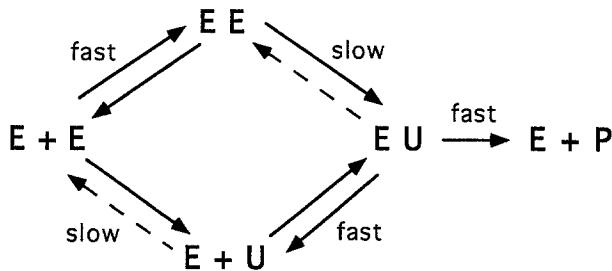


Fig. 3. Schematic representation of the overall thermal inactivation/autolysis reaction in NPs. E = intact enzyme; U = enzyme with at least one proteolytically susceptible loop unfolded to the extent that it can be cleaved; P = protein that has been irreversibly inactivated as a result of autolytic cleavage (which is rapidly followed by further proteolytic breakdown [114,115]). The two dashed arrows indicate that no equilibrium between E and U exists, because proteolysis is a much faster process than the preceding unfolding step.

thermolysin main chain in their preferred conformations, which were obtained from the WHAT IF relational database as described below.

The insertion around residue 27 in NP-ste (Fig. 1) was not modelled as database searches did not provide any clue how to do this. As a result, it is virtually impossible to use the NP-ste model to predict mutations in the direct 3D vicinity of residue 27.

In the second stage of the automated modelling procedure, unfavourable van der Waals overlap was removed by an automated debumping procedure, in which all freely rotatable side-chain torsion angles were iteratively rotated in steps of 120° or 180°. Then a short energy minimization, followed by four ps of MD and a final extensive energy minimization were carried out. The MD calculations were done in vacuo at 300 K, using the atomic partial charges incorporated in the GROMOS software package [128]. During the MD and EM runs constraints were applied to keep those atoms for which the coordinates were ‘known to be correct’, i.e. the residues that were identical in the X-ray structure and the model, within 0.25 Å of their original position.

The MD step is needed to get out of high-energy local minima that may contain errors that cannot be resolved by EM. Because of the in vacuo situation, the MD step only helps to refine the model in the interior parts, while the surface-located parts remain unrefined or may even deteriorate [57]. Unfortunately, MD simulations that include solvent molecules are complicated and require prohibitive amounts of computer time for molecules of the size of, for example, *Bacillus* NPs. The surface-located modelled side chains were therefore regularized, using the aforementioned position-specific rotamers.

The reliability of this strategy of model building by homology became apparent after the crystal structure of NP-cer was solved [72,73]. This NP has 73% sequence identity with thermolysin and its model was built before any information derived from crystals was available [129,130]. A comparison between the model and the structure (Fig. 4) gave an RMS atomic misplacement for all atoms in the complete protein of 1.26 Å. This error is dominated by a domain hinge-bending motion (see Fig. 4) and by errors in a limited set of side chains of surface-located residues which have large conformational freedom (like Glu, Lys, etc.) and consequently could not be modelled with any precision. Looking at individual domains of the NP, thus taking away the effect of the hinge-bending, the RMS atomic misplacement amounted to 1.11, 1.31, and 1.01 Å for the N-terminal, middle, and C-terminal domain, respectively. The middle domain exhibits the lowest accuracy probably because it largely consists of a long stretch of irregular structure, containing

residues with high crystallographic B-factors [71]. For nonexposed residues the accuracy of the model is much higher. Taking into account only residues having less than  $10 \text{ \AA}^2$  solvent-exposed area, the RMS atomic misplacement was reduced to  $0.68 \text{ \AA}$  for the complete protein and  $0.55$ ,  $0.77$  and  $0.55 \text{ \AA}$  for the individual domains, respectively.

The model building of NP-cer shows that local errors are only slightly larger than the errors expected for crystallographically determined structures. Apparently, the 73% sequence identity between thermolysin and NP-cer permits the construction of an acceptable model. Taking this as a reference it is obvious that the model of NP-ste (85% sequence identity with thermolysin) can be expected to be almost as good as a crystal structure for most of the molecule. Since NP-sub has only 47% sequence identity with thermolysin, its model is expected to be considerably less accurate than that of NP-cer. Apart from sequence differences, the presence of insertions and deletions contributes to this expected inaccuracy. It is noteworthy that the prediction of stabilizing mutations is more successful for NP-ste than for NP-sub.

Biochemical and genetical tools have been applied to get a better insight in the structure of NP-sub. Signor et al. [131] presented a model of NP-sub that was based partly on the results of limited proteolysis experiments, thus confirming the presence of a loose and protruding loop in the 215–225 region (Fig. 1). Analyses of autodigestion patterns of thermolysin and NP-sub [114,127] have shown that the two enzymes contain several identical autodigestion target sites, indicating that they have similar overall folds. Using a ‘model building by mutagenesis’ strategy, specific site-directed mutations were introduced and the effect of these mutations on thermostability was then taken as a clue for refining the model in the concerned parts of NP-sub. The results of this procedure will be described in detail elsewhere [132]; one case, concerning modelling of the region around residue 305, is described below.

#### *Modelling mutations*

It has recently been shown that the conformation of an amino acid side chain is determined mainly by the local backbone and much less by the 3D environment than was previously assumed [27,52]. Therefore position-specific rotamer searches are an important tool in modelling. In fact, it is our experience that position-specific appending of side chains gives the best results obtainable to date [27,52].

Position-specific rotamer distributions were obtained by searching the database for stretches of residues that had a similar local backbone conformation as the stretch around the position in the actual structure for which the rotamers were requested, and that had the same residue type at the middle position. Typically these stretches were five or seven residues long and the RMS difference between superposed  $\alpha$ -carbon atoms was kept at maximally  $0.5 \text{ \AA}$ . The two N- and C-terminal residues of proteins were excluded because they could not be the center of a stretch of residues with sufficient length needed for the distance geometry local structural-similarity search algorithm. Levitt recently described this method extensively [51].

Figure 5 shows the position-specific rotamers for histidine, phenylalanine and tyrosine at the second and the second to last position of a short helix. It is clearly seen that only one rotamer is allowed at the C-terminal end whereas the rotamer distribution is residue-dependent at the N-terminal end of the helix. It would go beyond the scope of this article to completely try to explain these effects, but Fig. 5 should suffice to stress the importance of position-specific rotamers for the prediction of mutant structures.

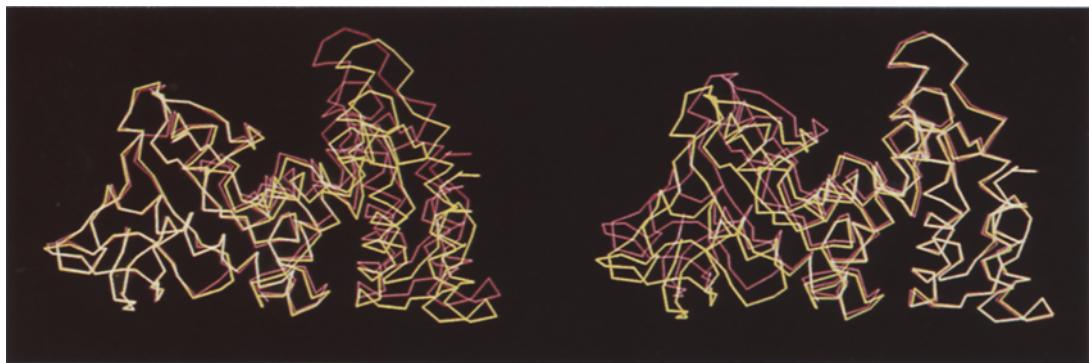


Fig. 4. Two possible superpositions of the modelled (red) and X-ray (yellow) structure of *B. cereus* neutral protease. In the left panel only the N-terminal domain was used for the superposition; in the right panel only the C-terminal subdomain was used. Only  $C_{\alpha}$  atoms are shown. The superpositions clearly show that the differences between the model and the X-ray structure result partly from hinge-bending between domains.

Singh and Thornton [53] determined directional preference patterns for residue-residue contacts. They found that many types of contacts show a strong asymmetric distribution. These contact patterns are incorporated in a relational manner in the WHAT IF database. This program also contains a module that compares the local contact patterns with the average contact patterns for similar residue-residue contacts in the database. This method has been described elsewhere [54], but it can be summarized as follows: If a residue-residue contact occurs often in the database such that the two residues have the same spatial orientation, then a high score is given. If a contact in the model molecule seems rather unique, either from a point of view of which residues make the contact, or from a point of view of directionality of the contact, a low score is given. This ‘quality control’ of local packing, based on a rather simple scoring scheme, has proven to be a powerful tool for the detection of abnormal structures and for the prediction of mutants [54].

The automatic mutant prediction module of WHAT IF is based on position-specific rotamer searches and quality control. It proceeds in three steps:

(1) Determine the position-specific rotamers for each of the 17 residue types that have a variable  $\chi^1$  at the position where a mutation is desired. Normally, one to three clusters are seen in  $\chi^1$  space, and per  $\chi^1$  angle only one or two preferred  $\chi^2$  angles.

(2) Determine the ‘average’ side-chain conformation in the  $\chi^1$  clusters with a high population density, mutate the residue in the molecule using this average rotamer and determine the packing quality. Of course rotamers that would give major Van der Waals clashes are excluded. Alanine and proline are modelled in only one conformation, and for glycine the existing side chain is simply removed. (In computo that is a very simple experiment)!

(3) A short MD run is performed on high-scoring mutants and the trajectory is visually analysed for anomalies. If nearby residues tend to move a lot then obviously conformational strain has been introduced, and re-evaluation (or rejection) of this mutant is required.

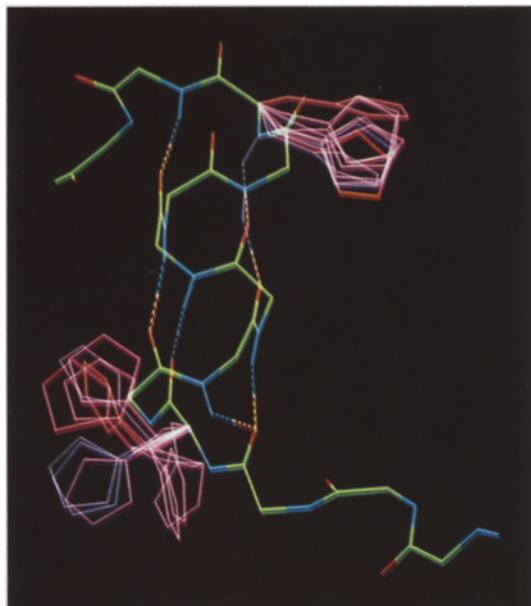
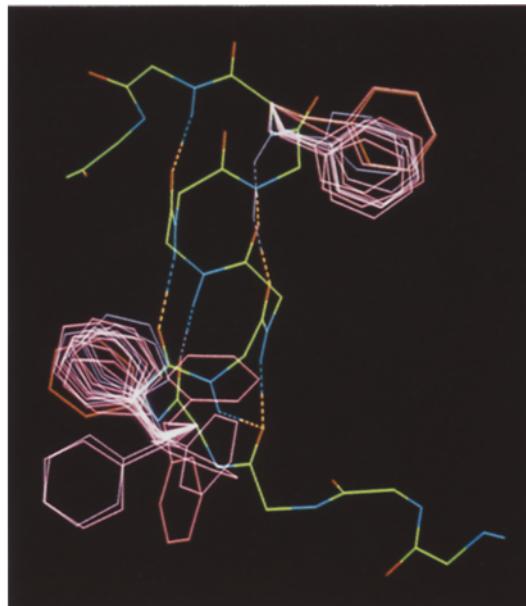
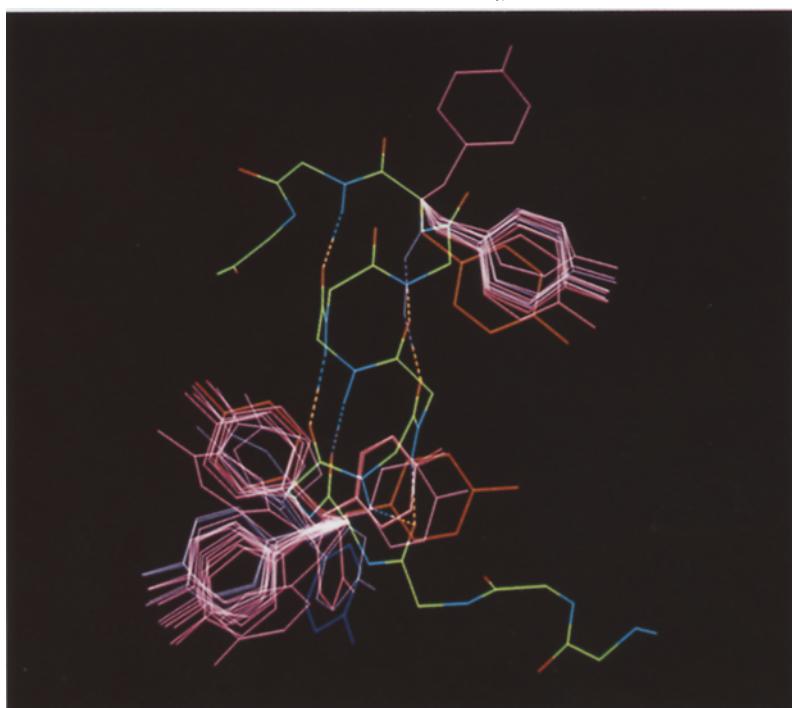
**A****B****C**

Fig. 5. Position-specific rotamers for (A) histidine; (B) phenylalanine; and (C) tyrosine at the second and the second-to-last position of a short helix. Details are described in the text.

It should be realized that this mutant prediction module only predicts potential mutant structures that look 'normal' when compared with existing protein structures. This means that the mutated protein will probably fold, but further selection criteria are needed to determine whether such a mutation will actually be stabilizing or not.

Table 2 shows the output of the automated module for two positions in NP-ste, which indicates the local packing quality for all possible amino acids at these positions. On the basis of the results shown in Table 2 two successful stabilizing mutations were designed, which are discussed below. It is still too early to say whether this method will always work, or that our first two trials were successful by accident. Some limitations of the method are discussed below ( $\text{Asn}^{238} \rightarrow \text{Leu}$  mutation).

## IDENTIFICATION, DESIGN AND ANALYSIS OF STABILIZING SITE-DIRECTED MUTATIONS

### *Design of mutations*

For the nonrandom identification of stabilizing site-directed mutations in proteins different approaches may be adopted, which are based to differing extents on knowledge of the interactions that stabilize proteins (see for example Refs. 133–136 for reviews).

TABLE 2  
PROGRAM OUTPUT BY THE AUTOMATED MUTANT PREDICTION MODULE IN WHAT IF FOR POSITIONS 163 AND 167 IN NP-ste

Position	163	167
Ala	+ 0.7	+ 0.6
Cys	+ 2.3	+ 2.2
Phe	—	—
Ile	+ 0.4	+ 0.3
Leu	+ 3.0	+ 1.9
Val	+ 0.7	+ 0.5
Trp	—	—
Tyr	—	—
Met	+ 0.8	+ 0.4
Asp	+ 3.4	+ 2.4
Glu	- 2.5	+ 2.7
His	—	+ 2.9
Lys	+ 1.8	+ 0.7
Asn	+ 3.4	+ 2.9
Gln	+ 0.8	+ 0.2
Arg	+ 1.3	+ 0.6
Ser	+ 4.5	+ 4.4
Thr	+ 2.7	+ 2.2
Gly	+ 0.7	- 0.4
Pro	—	0.0

Values indicate packing quality for the preferred rotamer, as described in the text. Missing values indicate residues that cannot be fitted in the model.

Statistical analyses of the primary structures of thermostable and thermolabile proteins have provided indications concerning preferential use of certain amino acids in thermostable enzymes [137–139]. It is tempting to use the ‘rules’ provided by these analyses as basis for a mutagenesis strategy, since they seem to have been applied successfully by nature and to be applicable without the necessity of structural analysis. This approach, however, is far too simplified, since the observed preferences may as well result from other evolutionary pressures (not in the least from genetic components like codon usage). Moreover, it is now clear that mutational effects do not depend on a certain ‘type’ of mutation, but are largely determined by the environment and the specific interactions of the mutated residue (e.g. [136,140–143]). The only widely tested stabilizing type of mutation that resulted from the statistical analyses and that seems to have a certain degree of general applicability is the Gly → Ala replacement. Lys → Arg mutations, forming a second ‘rule’ for stabilization, have recently been analysed and discussed extensively by Mrabet et al. [144].

In case homologous thermostable variants of the enzyme to be stabilized are available, sequence comparisons may point at residues critical for thermostability. For NPs this approach is very well applicable, because of the large amount of sequence data available. Examples include the stabilization of NP-sub by the introduction of the 248–257 loop [92], the Gly<sup>147</sup> → Ala mutation [94] and the Lys<sup>305</sup> → Ser mutation [91], and the stabilization of NP-ste by the Gly<sup>141</sup> → Ala mutation [86], the Thr<sup>63</sup> → Phe mutation [8] and the Ala<sup>69</sup> → Pro mutation (see below [95]). One of the very few stabilizing disulfide bridges in subtilisin was designed on the basis of the structure of a disulfide-containing thermostable homologous enzyme [145]. We call usage of this kind of information provided by nature ‘creative borrowing’.

Notwithstanding the successes based on them, sequence comparisons should be used with care: not every mutation in the direction of the more thermostable variant is stabilizing, and one should certainly include conserved residues in the search for possibly stabilizing mutations [89,94,146]. A striking example is the replacement of Ala<sup>167</sup> in NP-ste by Ser (see below). This mutation stabilizes the enzyme despite the fact that all thermostable NPs contain an Ala and all thermolabile NPs contain a Ser. This is a good illustration of the increasingly accepted viewpoint that, in engineering enzyme stability, every mutation has to be analysed structurally; sequence-based rules can create suggestions, but there are too many exceptions to use them blindly.

A more rational approach to the design of stabilizing mutations is based on visual and in computo inspection of the 3D structure of the protein to be stabilized: using general knowledge and experience concerning the stability of proteins, regions in which improvement of stabilizing interactions seems feasible are identified. In such regions, which for example contain cavities or unsatisfied hydrogen-bond donors or acceptors, possible mutations are studied using the modelling techniques described above.

Table 3 gives an overview of site-directed mutations in *Bacillus* NPs that were designed using the aforementioned approaches for finding and analysing stabilizing mutations. For the structural analysis of the mutations the modelling tools described above (position-specific rotamer searches, energy minimization, molecular dynamics, quality control, database searches) were used. The mutations relate to different types of stabilizing interactions in proteins. Their analysis and design are discussed in detail in the following sections.

#### *Hydrophobic interactions*

Hydrophobicity and high-quality packing of residues in the protein core are generally accepted

TABLE 3

EFFECTS OF SITE-DIRECTED MUTATIONS ON THE THERMOSTABILITY OF NEUTRAL PROTEASES FROM *B. SUBTILIS* (NP-sub) AND *B. STEAROTHERMOPHILUS* (NP-ste)

NP	Mutation	dT <sub>50</sub> (°C)	NP	Mutation	dT <sub>50</sub> (°C)
<b>Cavity filling/hydrophobic interactions</b>					
NP-ste	Leu <sup>144</sup> Phe	0	NP-ste	Ala <sup>69</sup> Pro	+ 5.6
	Met <sup>168</sup> Trp	+ 0.3			
	Ala <sup>240</sup> Val	0			
	Ala <sup>241</sup> Val	0			
	Leu <sup>284</sup> Trp	+ 0.4			
	Cys <sup>288</sup> Leu	+ 0.2			
	Cys <sup>288</sup> Ile	+ 0.4			
NP-sub	Phe <sup>168</sup> Trp	- 0.1			
	Ala <sup>240</sup> Val	- 2.3			
	Ala <sup>241</sup> Leu	<sup>a</sup>			
	Ala <sup>284</sup> Leu	<sup>a</sup>			
<b>Cavity creation/hydrophobic interactions</b>					
NP-ste	Met <sup>168</sup> Ala	+ 0.1			
	Leu <sup>284</sup> Ala	0			
	Cys <sup>288</sup> Ala	- 0.1			
<b>Gly → Ala mutations</b>					
NP-ste	Gly <sup>58</sup> Ala	n.d. <sup>b</sup>			
	Gly <sup>89</sup> Ala	+ 0.1			
	Gly <sup>109</sup> Ala	0			
	Gly <sup>135</sup> Ala	- 0.2			
	Gly <sup>136</sup> Ala	0			
	Gly <sup>141</sup> Ala	+ 0.6 <sup>b</sup>			
	Gly <sup>264</sup> Ala	+ 0.1			
NP-sub	Gly <sup>78</sup> Ala	labile <sup>c</sup>			
	Gly <sup>107</sup> Ala	labile <sup>c</sup>			
	Gly <sup>135</sup> Ala	- 1.9			
	Gly <sup>147</sup> Ala	stable <sup>d</sup>			
	Gly <sup>189</sup> Ala	stable <sup>d</sup>			

Residues are numbered according to their position in thermolysin, as indicated in Fig. 1. Mutational effects are given as dT<sub>50</sub> [115]. To determine this value purified enzymes were incubated at several temperatures at a concentration of approximately 0.1 μM at pH 5.0. After 30 min of incubation, residual proteolytic activities were determined and expressed as percentages of the initial activity. T<sub>50</sub> was defined as the temperature at which 50% of the initial activity is preserved. Mutational effects are expressed as dT<sub>50</sub>, being the difference in T<sub>50</sub> between the wild-type and the mutant enzyme. T<sub>50</sub>s of NP-sub and NP-ste were 55.3 °C and 68.5 °C, respectively. Error margins in T<sub>50</sub> were in the order of 0.5 °C; error margins in dT<sub>50</sub> were in the order of 0.2 °C. Some of these data have been described in earlier publications; see text for references. All mutant enzymes in this table, except the Gly<sup>141</sup>Ala mutant in NP-ste [214], exhibited specific activities towards casein that were similar to that of the corresponding wild-type enzyme.

<sup>a</sup> These enzymes were produced at very low levels and are expected to be very unstable.

<sup>b</sup> Mutation originally suggested by Imanaka et al. [86].

<sup>c</sup> Decrease in dT<sub>50</sub> of at least 6 °C.

<sup>d</sup> Stabilizing effects for these mutations were described by Margarit et al. [94]. Since these authors use a different stability assay no direct comparison of results is possible.

<sup>e</sup> Previously reported to be + 1.2 °C [89].

to be a major stabilizing factor in proteins [147–150]. The quality and density of hydrophobic packing is restricted by the fact that nature has only eight different hydrophobic amino acids available, each of them having only limited conformational space. Theoretical and site-directed mutagenesis studies have indicated that the hydrophobic core of a protein probably reflects a compromise between close packing and conformational strain [142,151–154] and internal cavities have been detected in many proteins [155].

Although the importance of hydrophobic interactions for the stability of a protein relative to, for example, hydrogen bonds is still a matter of debate (e.g. [156]), it is generally believed that improved hydrophobic packing will lead to improved protein stability. Site-directed mutagenesis studies have confirmed the importance of the hydrophobic core for protein stability [141,157,158], but examples of mutations that stabilize a protein via improvement of the hydrophobic core are scarce [152,159,160].

Cavities in the NPs were identified with the surface detection method of Voorintholt et al. [161], using a probe diameter of 1.5 Å [10] (the choice of the probe diameter is important and complicated as discussed by, for example, Eriksson et al. [158]). As shown in Table 3, NP-ste was quite insensitive to cavity-modulating mutations. The small stabilizing effects that are seen for some of the cavity-filling mutations indicate that the design of the mutations was in principle successful. The observed general insensitivity corroborates the view that local unfolding processes, presumably mainly involving surface-located areas, determine the thermostability of the NP. In contrast, cavity-filling mutations had mostly large destabilizing effects on NP-sub. Two factors that possibly contribute to this contrast are: (1) Because the model of NP-sub is less accurate, the mutations are not well designed and cause large destabilizing effects, such as conformational strain and steric clashes; and (2) The thermal unfolding processes that promote autolysis differ between NP-sub and NP-ste, involving more of the hydrophobic core in NP-sub. More extensive mutagenesis and thermal inactivation studies are needed to be able to discriminate between these two possibilities.

#### *Helix capping*

Statistical analysis of  $\alpha$ -helices has shown strong amino acid preferences for specific locations at the termini of  $\alpha$ -helices [162]. Residues capable of hydrogen-bond formation with backbone amides in the first turn of the helix are preferred at or near the so-called N-cap position [162,163]. There is a preference for negatively and positively charged residues in the first and last turn of the helix, respectively. Such charged, ‘helix-capping’ residues probably neutralize the  $\alpha$ -helix dipole, which arises from the regular alignment of peptide dipoles [164,165]. The peptide dipoles neutralize each other except in the first and last helical turn [166]. The remaining nonneutralized peptide dipoles constitute the macroscopic helix dipole that can be approximated by positioning half a positive charge at the N-terminal, and half a negative charge at the C-terminal helix end [165]. Site-directed mutagenesis experiments on several proteins have confirmed the importance of the interaction between charged residues and the helix dipole [91,167–172]. Series of mutations at N-cap residues in Barnase and T4 lysozyme have given insight in the stabilizing effects of both hydrogen-bond formation and charge-dipole interactions [169,173].

The applicability of helix-capping mutations was tested by replacing lysine residues in surface-located N-terminal turns of  $\alpha$ -helices of NP-sub by Ser and Asp (Table 3 [91]). The feasibility of such replacements was estimated by simply looking at the possibility to put the introduced side

chains in their most preferred rotamers, without the introduction of negative side effects. It should be noted that negative side effects of a mutation include both the introduction of unfavourable interactions and the removal of beneficial interactions. The effect of removing Lys<sup>282</sup> was hard to predict because of the nearby presence of a negatively charged group (Asp<sup>283</sup>), which could interact with both the helix dipole and with the positive charge of the lysine. The K282S mutation turned out to be the only destabilizing one, out of the five tested. The mutations at positions 262 and 305, which in the modelling process did not show negative side effects, stabilized the enzyme.

The fact that helix ends most often occur at the protein surface makes helix capping one of the better strategies for protease stabilization.

#### *Salt bridges*

The importance of salt bridges for protein stability has been suggested as early as 1975 [174]. Studies on the effects of salt-bridge formation on the stability of synthetic  $\alpha$ -helical peptides gave results that are in favour of this suggestion, although some conflicting data exist [175–177]. Recent protein engineering studies have provided evidence that the contribution of salt bridges to protein stability is generally low [170,171,178–180], although for nonsurface-located bridges occasionally large stabilizing effects have been observed [181]. Application of salt-bridge engineering to subtilisin BPN' showed similar results: out of 10 engineered salt bridges, only one had a (small) stabilizing effect [182].

#### *Hydrogen bonding*

The contribution of hydrogen bonding to overall protein stability is somewhat controversial, since hydrogen-bond formation can occur in both the folded and the unfolded protein, and in the transition state of (un)folding [149,156,183]. The importance of specific hydrogen bonds for protein stability has been shown by site-directed mutagenesis experiments [184,185]. Mutations aimed at improving hydrogen-bonding networks, for example by satisfying unsatisfied hydrogen-bond donors and acceptors in the protein interior, were shown to have stabilizing effects in subtilisin [186,187] and NP-ste (Table 3 [88–90]).

A search for buried potential hydrogen-bond donors and acceptors that are not involved in hydrogen bonds (using the parameter relation row option in WHAT IF [188]) focused attention on Asn<sup>238</sup>. This polar residue, which is conserved in all neutral proteases, is located in a mostly apolar environment just beneath the surface of NP-ste. The hydrogen-bonding requirements of its polar side chain are only marginally satisfied and the residue is not important for satisfying the hydrogen-bonding requirements of other residues [88]. Quality control of position 238 showed preferences for polar residues, Asn being the best one. Here limitations of the quality-control procedure emerge: it does not take into account negative factors, nor does it take into account differences between the folded and the (locally) unfolded state of the protein. Asn comes out as a favourable residue because of a hydrogen bond between its Nδ2 and a neighbouring Asp-Oδ1, which is already hydrogen bonded to another hydrogen-bond donor in the molecule. The clear negative effect of burying a completely unsatisfied hydrogen-bond acceptor (Asn-Oδ1) at the same time is not counted in this outcome. In other words: Asn<sup>241</sup> → Asn<sup>238</sup>-Nδ2 scores positive, whereas Asn<sup>241</sup> → Asn<sup>238</sup>-Oδ1 scores zero. Similar reasoning applies to the other polar residues suggested by the quality-control analysis. Taking into account the artefacts emerging from quality control, we decided to replace Asn by an apolar residue anyway. Leucine was chosen for

three reasons. First, it is the best of all apolar residues according to the quality-control method. Second, it is isosteric with Asn. Third, Leu acquires optimal side-chain torsion angles when it is placed in the NP-ste model such that it maximally overlaps with the original Asn. As expected on the basis of these modelling studies, the Asn<sup>238</sup> → Leu mutation stabilized NP-ste (Table 3).

The automated mutant prediction module in WHAT IF was used to analyse the possibilities for mutations in a small arbitrarily chosen helical part of NP-ste (residues 161–170). At four positions the program suggested possible stabilizing mutations. At position 162 a Gly → Ala mutation was suggested. Because of the special entropical effects of Gly → Ala mutations on the unfolded state of a protein [189], we decided not to make this mutation; not because it would not stabilize, but because if it would, we would not know why. At position 164 (Ile in the wild-type) Glu, Asp and Ser are suggested. However, we could not envisage that one of these mutations would stabilize the protein. The program strongly suggests substitution of Ala<sup>163</sup> and Ala<sup>167</sup> by Ser (see Table 2). At first glance these two mutations seemed to improve hydrogen bonding in NP-ste and they were therefore studied in more detail.

Inspection of the model of NP-ste suggested the following effects for the A163S mutation [89]: (1) The number of hydrogen bonds is increased by one in both the folded and the unfolded state. Thus, in this respect the folding/unfolding equilibrium is unchanged; (2) The number of intraprotein hydrogen bonds is increased, thus probably increasing the activation free energy of unfolding; (3) A buried water molecule is replaced by the hydroxyl group of the Ser and thus transferred from the protein interior to the solvent; consequently a favourable increase in entropy is obtained for the folded state of the mutated protein. From these considerations it was anticipated that the Ala<sup>163</sup> → Ser mutation would have a stabilizing effect.

Figure 6 gives an overview of the modelling procedure that was used to evaluate the A167S mutation. The Ser side chain was initially modelled, using the standard rotamer search procedure. As shown in Fig. 6, the subsequent EM-MD procedure showed the reason for the stabilizing effect of this mutation: within the first 0.3 ps of an MD run, the side chain of Asn<sup>238</sup> reoriented itself to a position in which its unsatisfied Oδ1 (see also above) and the hydroxyl group of Ser<sup>167</sup> can form a nice hydrogen bond. The hydrogen-bond requirements of other groups were hardly affected by the reorientations during the MD run. The positive effect of the extra hydrogen bond is partly offset by the fact that the introduced hydroxyl group of Ser<sup>167</sup> has hydrogen-bond requirements that are only partly satisfied in the folded enzyme. Still, a stabilizing effect was expected for the A167S mutation, because of a stabilization of the folded state relative to the transition state of (local) unfolding [90].

As shown in Table 3, the design procedures for the improvement of hydrogen-bonding networks in NP-ste indeed resulted in the correct prediction of stabilizing mutations. It should be noted that these mutations concern conserved residues (Fig. 1) that would definitely not have been selected for mutagenesis on the basis of sequence comparisons. The results stress the value of procedures involving a relational protein-structure database and packing quality-control procedures, for designing stabilizing mutations.

#### *Binding of calcium ions*

Binding of calcium ions plays an important role in stabilizing several enzymes including NPs [81,93] and subtilisin [80,190]. In proteases calcium ions contribute to stability by protecting the corresponding ion-binding loop from local unfolding and subsequent autolysis [112,115,122]. In

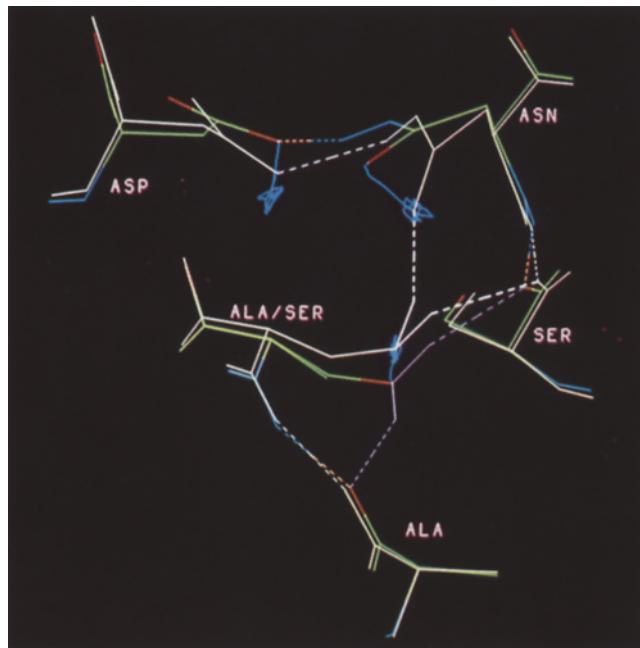


Fig. 6. The environment of residue 167 (Ala/Ser) in the neutral protease of *B. stearothermophilus*, before and after MD. All relevant residues and hydrogen bonds (dashed lines) are shown. Hydrogens involved in hydrogen bonds are drawn as solid lines. A superposition of the wild-type structure (Ala) and the initial model of the mutant (Ser) is shown in full colour (carbons green, oxygens red, nitrogens blue; possible positions for the Ser<sup>167</sup> hydrogens are indicated by purple lines). The model of the mutant after MD is shown in white. The curved blue lines show the 2-ps trajectory of some relevant atoms.

accordance with this, a mutant of human lysozyme that was stabilized after introducing a calcium binding site appeared to be more resistant to protease digestion than the wild-type enzyme [191].

Thermostable NPs bind more calcium ions than thermolabile NPs. Toma et al. [93] have shown that the thermolabile NP-sub can be stabilized by replacing a non-calcium-binding loop by the homologous calcium-binding loop from thermolysin. Further protein engineering studies on the role of calcium binding in NP stability are currently being conducted in several laboratories.

#### *Gly → Ala → Pro mutations*

Site-directed mutagenesis experiments have shown that Gly → Ala mutations are often successful in stabilizing a protein [86, 94, 143, 189, 192, 193]. The stabilizing mechanism of these mutations is, however, a matter of debate and different authors have attributed the effects to different factors. From a theoretical point of view, Gly → Ala mutations have a clear stabilizing entropic effect, caused by the fact that the backbone around glycine has greater rotational freedom than any other residue in the unfolded state. Thus, Gly → Ala mutations reduce the entropy that is gained upon unfolding [189]. Most stabilizing Gly → Ala mutations described so far concern residues which are located in  $\alpha$ -helices. The stabilizing effects could therefore also be the result of positive effects on helix stability [143, 194, 195]. In general, the introduction of alanines in most positions in  $\alpha$ -helices seems to have a positive effect on protein stability [143, 196, 197]. A third possible stabilizing effect of Gly → Ala mutations could result from changes in hydrophobic

interactions within the protein and changes in protein-solvent interactions [143,193]. Serrano et al. [143] have clearly shown that the effect of replacing Gly by Ala in an  $\alpha$ -helix is strongly dependent on the precise context of the mutated residue (position in the helix; solvent accessibility).

In NP-sub and NP-ste several Gly  $\rightarrow$  Ala mutations have been made (Table 3). The three mutations with clear stabilizing effects contain residues that are located in an  $\alpha$ -helix (141,147) and/or are shielded from the solvent and involved in hydrophobic interactions within the protein (141,147,189). Structural analysis showed that in all three cases the alanine side chains could be accommodated without the introduction of Van der Waals overlap. The Gly  $\rightarrow$  Ala mutations that have no effect or even a destabilizing effect on thermostability concern solvent-exposed positions or positions at which the alanine C $\beta$  introduces some steric hindrance. At these positions the entropic effect seems to be the only positive aspect of Gly  $\rightarrow$  Ala exchanges. Apparently, this effect is either not present or it is completely offset by the negative side effects.

There are indications that proline residues contribute to enzyme thermostability [189,198–200]. This contribution could be explained in terms of entropic effects (or chain flexibility) since proline is conformationally more restricted in the unfolded state than any other amino acid [189]. Other consequences of the introduction of proline residues can be positive or negative changes in hydrophobic contacts. Since proline residues are very restricted with respect to their  $\phi$ - $\psi$  angles, their introduction can easily cause conformational strain in the backbone. We introduced one Ala  $\rightarrow$  Pro mutation at the partly solvent-exposed position 69 in NP-ste [Table 3], which had a drastic stabilizing effect. This effect can be attributed to a combination of improved hydrophobic packing just underneath the protein surface and to reduced flexibility of the surface-located environment of the mutated residue. A large set of Xxx  $\rightarrow$  Pro mutations, aimed at improving thermostability, was made in Subtilisin NOVO. Both stabilizing and destabilizing effects were observed. These experiments did not reveal any physical parameter that can be correlated to the effect of introduction of a Pro residue on thermostability [201].

#### *Disulfide bridges*

Disulfide bridges are also considered to contribute to stability by decreasing the entropy of the unfolded state [202,203]. The production of active enzymes having correctly formed disulfide bonds, introduced by site-directed mutagenesis, has been reported. The effects of the engineered disulfides on protein stability were mixed [117,133,204]. Observed destabilizing effects have been attributed to conformational strain caused by the disulfide formation and to the negative effects of replacing wild-type residues with the cysteines required for disulfide formation [116,133,204]. Introduction of disulfide bridges in nonproteolytic enzymes was successful in several cases [204–206]. Mutants of T4 lysozyme containing multiple engineered disulfide bonds exhibited dramatic increases in thermostability [207].

Introduction of designed disulfide bridges was particularly unsuccessful in proteases ([113,116,117,208,209]; see Ref. 145 for a successful mutation). This lack of success can partly be attributed to deleterious effects of the introduction of the individual cysteines [113,117,209]. In addition, the specific mode of irreversible thermal inactivation in protease may play a role [113,116,117,209].

#### *Design on the basis of sequence comparisons*

In case of NPs, a comparison of the known naturally occurring homologous thermolabile and

thermostable variants has been used successfully for the design of stabilizing mutations. Indeed, some of the mutations described above also occur in nature. A striking example of this approach was recently published by Van den Burg et al. [8], who cloned and sequenced the gene encoding the NP of *B. caldolyticus* (NP-cal) and characterized the corresponding protein. Compared to NP-ste, NP-cal appeared to carry three mutations (positions 4, 56 and 63), which made this enzyme 8.2 °C more thermostable. By site-directed mutagenesis in NP-ste, the effects of the individual mutations were determined as shown in Table 3. Clearly, nature has provided us with some of the best stabilizing mutations in NP-ste. It is somewhat discouraging that these mutations would most probably never have been selected on the basis of modelling studies using the techniques available today. A detailed description of this comparative approach is beyond the scope of this review.

## MODEL BUILDING BY MUTAGENESIS

Several techniques exist to model loops if no, or insufficient, homology is available [35–41]. All these methods suffer from one serious problem. They generate a large number of alternative conformations, but normally only one conformation is used for the final model. One therefore can only be sure that the modelled loop has a plausible conformation, but the chance that the chosen alternative is the correct one is much smaller than in other steps of the modelling procedure. The problem of modelling without homology occurs always near insertions or deletions, but sometimes also in areas where there is no sequence homology for many residues in a row. In case of NP-sub, several insertions and deletions had to be modelled. Some cases were essentially impossible to model with any degree of confidence. In those cases where the number of plausible alternatives was limited, we have designed mutations that should distinguish between the potential conformations.

One example is the deletion near the N-terminal end of the C-terminal helix (residues 300 and 301). After an extensive study of the structure of thermolysin, it was concluded that there were only two ways that this deletion could be modelled. These two alternatives are shown in Fig. 7. In alternative A (Fig. 7A) the N-terminal turn of the helix is modelled as in thermolysin, and the deletion is placed in the loop that precedes the helix. This loop connects the N-terminal end of the C-terminal helix with the C-terminal end of the preceding helix. In alternative B (Fig. 7B) the first turn of the C-terminal helix (as present in thermolysin) is unwound to keep the same number of residues in the loop between the two helices as in thermolysin. In conformation A the solvent-exposed lysine at position 305 is far away from the N-terminal end of the helix, and thus far away from the positive pole of the helix dipole. Its positively charged amino group is at salt-bridge-forming distance of Asp<sup>302</sup>. In solution B, the lysine is much closer to the dipole and much further away from the side chain of Asp<sup>302</sup>. Replacing the positively charged Lys by an uncharged (Ser) or a negatively charged (Asp) residue therefore seems favourable if conformation B would be valid, whereas it seems unfavourable if conformation A would be valid. As described above the Lys<sup>305</sup> → Ser (+ 1.0 °C) and the Lys<sup>305</sup> → Asp (+ 1.2 °C) mutations have significant positive effects on the thermostability of NP-sub, indicating clearly that the conformation shown in Fig. 7B is the better model for the real structure of NP-sub in this area.

Many such mutations were made; some even included rather elaborate combinations of loop transplantations and point mutations [132]. However, in practice we consider every correct

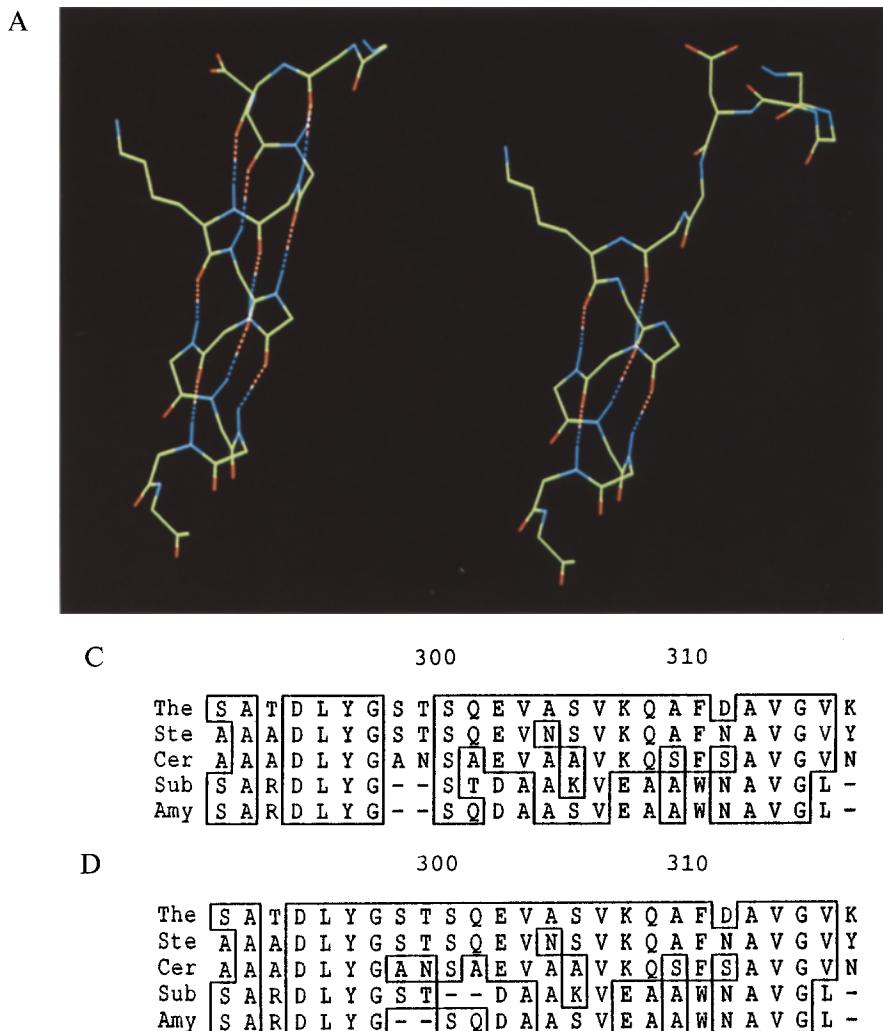


Fig. 7. Two possible structures and alignments for the environment of Asp<sup>302</sup> and Lys<sup>305</sup> in the neutral protease of *B. subtilis*. See text for details. The left conformation, A, belongs to alignment C; B belongs to alignment D.

prediction of the effects of a point mutation to be a confirmation of the correctness of the method. Incorrect predictions, on the other hand, are either indicative of a model error, or of a lack of our understanding of the structural rules that govern thermostability.

#### THERMAL UNFOLDING OF NEUTRAL PROTEASES

The irreversible denaturation of neutral proteases results from local unfolding of surface areas (normally called surface loops, although the local unfolding processes may involve more extensive regions of the protein) followed by autolysis [9,115]. Let us now follow through a train of thought. Let us assume that five loops, far apart in the structure, can locally unfold to the extent that they become susceptible towards autolysis. The fact that at temperatures around  $T_{50}$  only

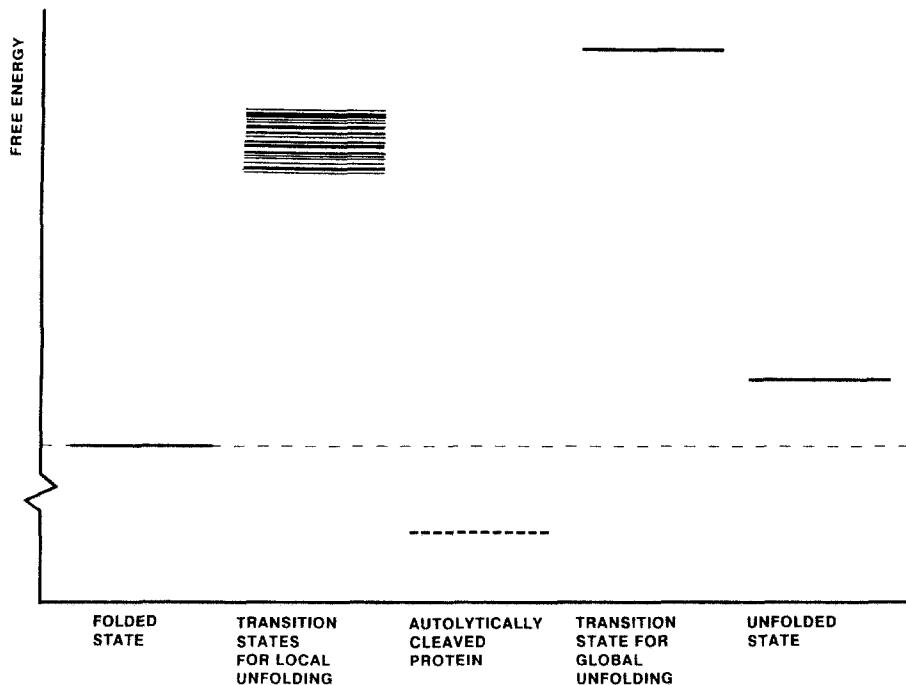


Fig. 8. Schematic free-energy diagram for the inactivation reaction of a neutral protease at physiological temperature. Each transition state of local unfolding represents the activation free energy of an unfolding pathway that results in an autolytically susceptible, locally unfolded molecule. The hypothetical transition state for global unfolding and the free energy of the unfolded state are shown for reference.

half of the NP molecules gets cleaved in a 30-min incubation period (details of the stability assay are described in the legend to Table 3 and in previous publications [9,115]) shows that the rate-limiting unfolding steps are slow ( $\sim 10^{-4} \text{ s}^{-1}$ ). Each of the five loops gives rise to a certain number of cleavages per second, and thus contributes to a certain extent to the denaturation rate. Now let us assume that we have two mutations that involve residues that have no direct contact and that each of them can stabilize the same loop with say 2 kcal/mol. The local unfolding rate of a loop decreases dramatically by such a large stabilization. That means that one of the two mutations is already sufficient to make the loop virtually uncleavable in the time period of the stability assay. The rate of denaturation will now be determined by cleavage in the other four loops. The second mutation, which also stabilizes by 2 Kcal/mol, cannot make the already stabilized loop less cleavable, and thus no extra stabilization will be observed after adding this second mutation. This 'enough = enough' principle has indeed been observed several times [9] and shows that local unfolding is the rate-limiting step in NP inactivation, because if the autolysis would have to be preceded by global unfolding, then the effects of two mutations that do not contact each other directly should in principle be additive [210].

A hypothetical schematic energy diagram for the thermal denaturation of NPs is depicted in Fig. 8. The diagram reflects the existence of a large number of loops that can unfold. The actual number and nature of these loops is of great importance for analysing and engineering NP stability. The possibility exists that one loop unfolds much faster than all others, or, in energetic

terms, that the activation energy occurring in one of the local unfolding pathways is much lower than in all the others. In that case it would be expected that only mutations aimed at the stability of this so-called ‘weakest link’ have significant effects on the thermostability. A rather stable loop will not be cleaved and mutations that stabilize this loop will thus have no effect on the observed  $T_{50}$ . Also, after a destabilizing mutation such a stable loop may not yet have become cleavable to the extent that measurable effects on  $T_{50}$  result.

Figure 9 shows the  $\text{C}\alpha$  tracing of the model of NP-ste coloured as function of the absolute value of the  $dT_{50}$  of mutations. In this figure one very sensitive region (yellow) and a few moderately sensitive regions (red) can be seen. The 50–70 region in the N-terminal domain is clearly a very weak link in which point mutations have effects ranging from  $-16$  to  $+7.0\text{ }^\circ\text{C}$  [95,96]. An extensive network of atomic contacts exists between this area and the underlying N-terminal  $\beta$ -hairpin (residues 3–18) in which until now one mutation having a considerable effect on thermostability (A4T,  $dT_{50} = +1.8\text{ }^\circ\text{C}$ ) has been constructed. We are presently analysing a large number of mutations aimed at determining whether the N-terminal  $\beta$ -hairpin is stabilizing the 60–70 loop or vice versa. The residues 66–70 form a somewhat irregular N-terminal turn of an  $\alpha$ -helix, and instability of this helix may therefore also play a role. The important conclusion so far is that NP-ste has a clear ‘weak link’. This became even more clear after combining site-directed mutations in this area, which gave an increase in stability of over  $10\text{ }^\circ\text{C}$  [146]. Remarkably, a second sensitive area is located close to the active site of NP-ste, a region which, by its nature, is supposed to be relatively flexible [205].

Using Eyring’s steady-state theory we calculated that the free energy of activation for the overall thermal denaturation process at  $T = T_{50}$  is approximately 20 Kcal/mol [115]. This value is in the same range as those estimated for the free energy of activation for global unfolding of globular proteins [108,183,211]. So, in the state in which the NP is locally unfolded to the extent that autolysis can occur, the protein is energetically already very close to global unfolding, probably resembling the transition state for global unfolding as described in the introduction of this paper.

It should be noted that the importance of local unfolding as suggested here does not contradict the general view that protein unfolding is a highly cooperative process that can often be described by a relatively simple two-state model. As indicated in the hypothetical unfolding diagram of Fig. 10, in the case of NPs the cooperative unfolding process is interrupted by autolysis (at  $T_{50}$ ) before global unfolding (at  $T_m$ ;  $T_m > T_{50}$ ) occurs. Our views on the thermal inactivation mechanism of NPs imply the presence of locally unfolded intermediates in the unfolding pathway that live long enough to be prone to autolysis.

The presence of a ‘weak link’ gives important information on the early steps in NP-ste unfolding. Apparently unfolding starts in the N-terminal domain, perhaps in the  $\beta$ -pleated structure involving the N-terminal  $\beta$ -hairpin. This observation is in accordance with earlier observations on the unfolding and stability of thermolysin [212]. More insight in this matter will result from further mutagenesis studies and an extensive MD simulation in solvent, which are currently being conducted in our laboratories.

## CONCLUDING REMARKS

The above results show that protein models built by homology can be sufficient for the identifi-

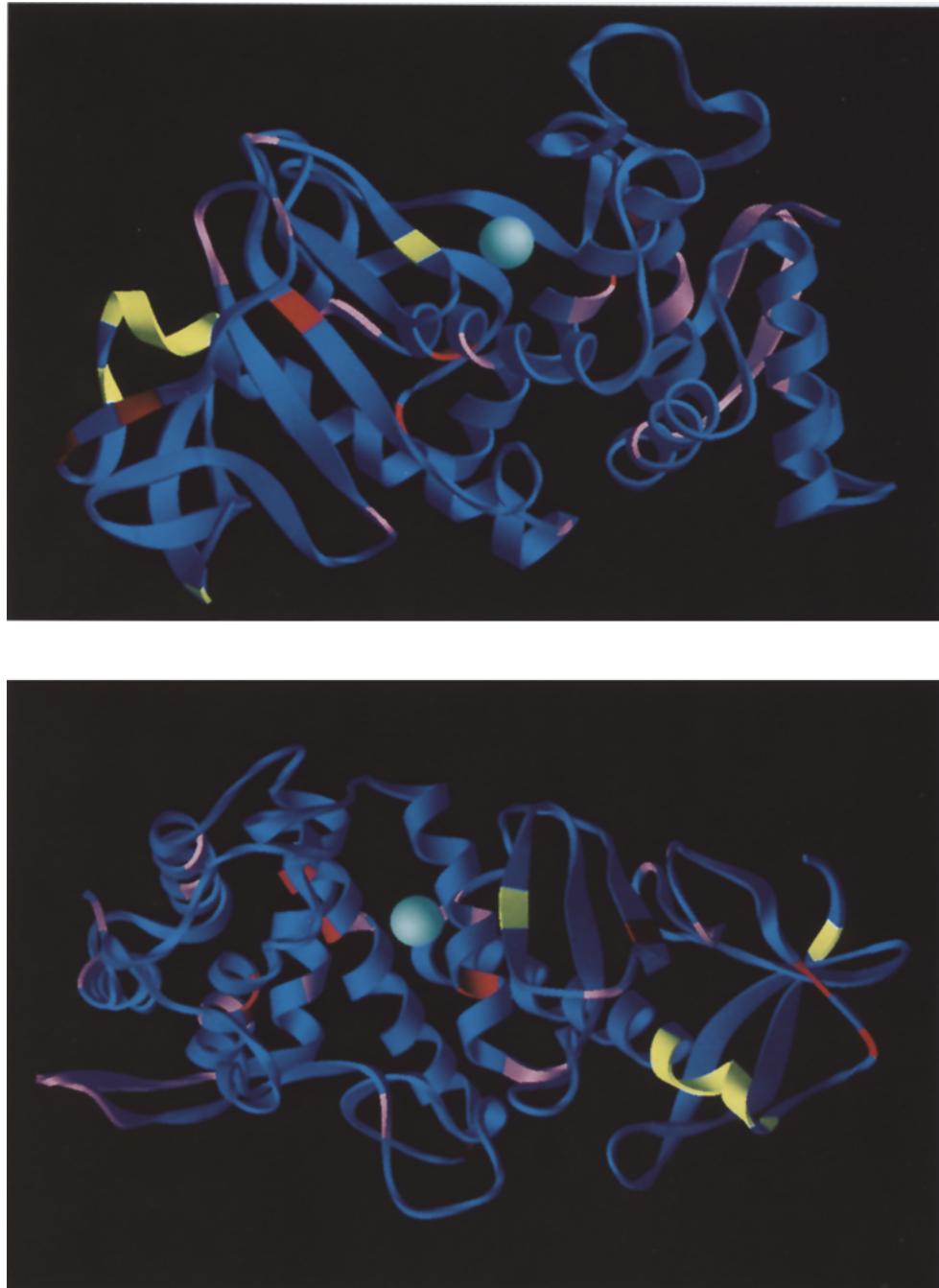


Fig. 9. Two ribbon representations [215] of the neutral protease of *B. stearothermophilus*, indicating regions in which site-directed mutations have large effects on thermostability. Blue: no mutations made; purple: absolute mutational effect smaller than 0.5 °C; red: absolute mutational effect between 0.5 and 2.0 °C; yellow: absolute mutational effect over 2.0 °C. Data from Table 3 and Refs. 8–10, 88–90, 95, 96.

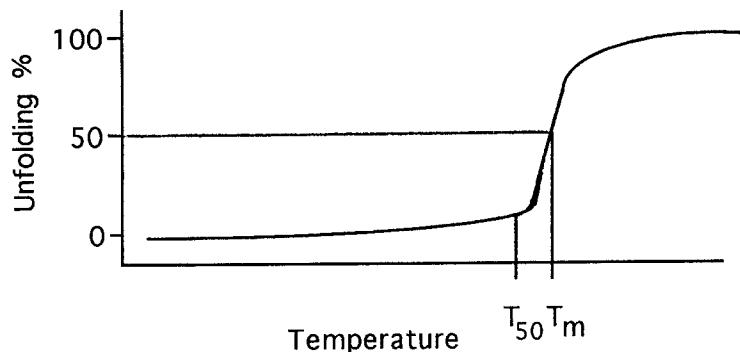


Fig. 10. Hypothetical diagram for (cooperative) protein unfolding.

cation and quantitative analysis of stabilizing site-directed mutations. In NP-sub mutations were less successful than in NP-ste, indicating the existence of larger model building errors in the former enzyme. Thus, high sequence homologies (probably over 75%) seem to be needed in order to obtain models of sufficient quality for reliable quantitative mutant predictions. Model building by mutagenesis may help to improve presumably inaccurate models. However, this strategy is limited to situations where the uncertainties in the model involve only a few alternative conformations. Also, modelling by mutagenesis is much more laborious than in computo modelling.

The use of position-specific rotamers for side-chain modelling has been shown to be a powerful tool for model building and for mutant prediction. The success of their use in the present study corroborates earlier observations that indicated their value in model building [52].

The feasibility of predicting stabilizing mutations in NPs, even in a fully automated fashion, was clearly shown. However, our understanding of protein stability has not yet progressed far enough to be able to beat nature systematically. The best stabilizing mutations were creatively borrowed from homologous enzymes. Since the use of information that is hidden in existing proteins can be quite successful in mutation-prediction work, the design of tools that allow faster and better creative borrowing is of great importance. Property profile databases and packing quality control are among the examples of tools that allow us to extract information from the large body of known protein structures [213].

Most of the mutations that were selected on the basis of molecular modelling indeed stabilized the NPs, but their effects were small as compared with some of the creatively borrowed mutations such as  $\text{Thr}^{63} \rightarrow \text{Phe}$  in NP-ste. As described above, the magnitude of mutational effects in NPs is strongly determined by the location of the mutation. The fact that several mutational effects were small most probably reflects the fact that they were introduced in relatively insensitive ('nonweak') regions of the protein. Recently, several designed mutations were made in the sensitive 50–70 region in NP-ste which gave much larger stabilizing effects [95].

The thermostability of NPs is kinetically controlled by local unfolding of surface loops that become susceptible to autolysis. This specific mechanism has important consequences for the design and analysis of site-directed mutations. Because of this mechanism, the effects of site-directed mutations on the thermostability of NPs can be used to monitor the early steps of unfolding. An alanine scan could be a possible tool to detect mutationally sensitive areas that presumably unfold early. There is no reason to expect the early unfolding steps in NPs to be

fundamentally different from those in other proteins, so the studies on early unfolding of NPs may provide some general insight in protein unfolding.

The effect of a mutation is determined by its environment in the molecule. Without thorough analysis of the local structure it is not possible to predict the effects of a mutation on the structure, function and stability of a protein.

## ACKNOWLEDGEMENTS

We thank the many colleagues who contributed to this work for stimulating and critical discussions, providing computer software, reading parts of the manuscript, and experimental work. Special thanks go to Chris Sander and Gerard Venema for their continuous support of the work.

## REFERENCES

- 1 Giege, R. and Mikol, V., *Trends Biotechnol.*, 7 (1989) 277.
- 2 Swindells, M.B. and Thornton, J.M., *Curr. Opin. Struct. Biol.*, 1 (1991) 219.
- 3 Bajaj, M. and Blundell, T.L., *Annu. Rev. Biophys. Bioeng.*, 13 (1984) 453.
- 4 Bashford, D., Chotia, C. and Lesk, A.M., *J. Mol. Biol.*, 196 (1987) 199.
- 5 Chotia, C. and Lesk, A.M., *EMBO J.*, 5 (1986) 823.
- 6 Sander, C. and Schneider, R., *Proteins*, 9 (1991) 56.
- 7 Blundell, T.L., Sibanda, B.L., Sternberg, M.J.E. and Thornton, J.M., *Nature*, 326 (1987) 347.
- 8 Van den Burg, B., Enequist, H.G., Van der Haar, M.E., Eijsink, V.G.H., Stulp, B.K. and Venema, G., *J. Bacteriol.*, 173 (1991) 4107.
- 9 Eijsink, V.G.H., Vriend, G., Van der Vinne, B., Hazes, B., Van den Burg, B. and Venema, G., *Proteins*, 14 (1992) 224.
- 10 Eijsink, V.G.H., Dijkstra, B.W., Vriend, G., Van der Zee, J.R., Veltman, O.R., Van der Vinne, B., Van den Burg, B., Kempe, S. and Venema, G., *Protein Eng.*, 5 (1992) 421.
- 11 Fontana, A., *Biophys. Chem.*, 29 (1988) 181.
- 12 Taylor, W.T., *J. Mol. Biol.*, 188 (1986) 233.
- 13 Novotny, J., Rashin, A.A. and Brucolieri, R.E., *Proteins*, 4 (1988) 19.
- 14 Overington, J., Donnelly, D., Johnson, M.S., Sali, A. and Blundell, T.L., *Protein Sci.*, 1 (1992) 216.
- 15 Crippen, G.M., *Biochemistry*, 30 (1990) 4232.
- 16 Sippl, M.J. and Weitckus, S., *Proteins*, 13 (1992) 258.
- 17 Luthy, R., McLachlan, A.D. and Eisenberg, D., *Proteins*, 10 (1991) 229.
- 18 Bowie, J.B., Luthy, R. and Eisenberg, D., *Science*, 253 (1991) 164.
- 19 Luthy, R., Bowie, J.U. and Eisenberg, D., *Nature*, 356 (1992) 83.
- 20 Jones, D.T., Taylor, W.R. and Thornton, J.M., *Nature*, 358 (1992) 86.
- 21 Scharf, M., Ph.D. Thesis, University of Heidelberg, 1989.
- 22 Thornton, J.M., Flores, T.P., Jones, D.J. and Swindells, M.B., *Nature*, 354 (1991) 105.
- 23 McGregor, M.J., Islam, S.A. and Sternberg, M.J.E., *J. Mol. Biol.*, 198 (1987) 295.
- 24 Ponder, J.W. and Richards, F.M., *J. Mol. Biol.*, 193 (1987) 775.
- 25 Read, L.S. and Thornton, J.M., *Proteins*, 5 (1989) 170.
- 26 Summers, N.L. and Karplus, M., *J. Mol. Biol.*, 210 (1989) 785.
- 27 Eisenmenger, F., Argos, P. and Abagyan, R., *J. Mol. Biol.*, 231 (1993) 849.
- 28 Jones, T.A. and Thirup, S., *EMBO J.*, 5 (1986) 819.
- 29 Schomburg, D. and Reichelt, J., *J. Mol. Graphics*, 6 (1988) 161.
- 30 Dayringer, H.E., Tramontano, A. and Fletterick, R.J., *J. Mol. Graphics*, 4 (1986) 82.
- 31 Vriend, G., *J. Mol. Graphics*, 8 (1990) 52.
- 32 Jones, T.A., *J. Appl. Cryst.*, 11 (1978) 268.

- 33 Hubbard, R.E., In Fletterick, R.J. and Zoller, M. (Eds), Computer Graphics and Molecular Modelling, Cold Spring Harbor, 1986, pp. 9–12.
- 34 Jones, T.A., Zou, J.Y., Cowan, S.W. and Kjelgaard, M., *Acta Cryst.*, A47 (1991) 110.
- 35 Moult, J. and James, M.N.G., *Proteins*, 1 (1986) 146.
- 36 Brucolieri, R.E. and Karplus, M., *Biopolymers*, 26 (1987) 137.
- 37 Fine, R.M., Wang, P.S., Shenkin, D.L., Yarmush, D.L. and Levinthal, C., *Proteins*, 1 (1986) 342.
- 38 Havel, T.F. and Snow, M.E., *J. Mol. Biol.*, 217 (1991) 1.
- 39 Sippl, M.J., Hendlich, M. and Lackner, P., *Protein Sci.*, 1 (1992) 625.
- 40 Simon, I., Glasser, L. and Scheraga, H.A., *Proc. Natl. Acad. Sci. U.S.A.*, 88 (1991) 3661.
- 41 Ripoll, D.R. and Scheraga, H.A., *Biopolymers*, 30 (1990) 165.
- 42 Lesk, A.M. and Chothia, C., *J. Mol. Biol.*, 136 (1980) 225.
- 43 Chothia, C. and Lesk, A.M., *J. Mol. Biol.*, 182 (1985) 151.
- 44 Chothia, C. and Lesk, A.M., *Trends Biochem. Sci.*, 190 (1985) 116.
- 45 Richardson, J.S., In Anfinsen, C.B., Edsall, J.T. and Richards, R.M. (Eds), Advances in Protein Chemistry, Vol. 34, Academic Press, pp. 167–339.
- 46 Ashida, T., Tsunogae, Y., Tanaka, I. and Yamane, T., *Acta Cryst.*, B43 (1987) 212.
- 47 Islam, S.A. and Sternberg, M.J.E., *Protein Eng.*, 2 (1989) 431.
- 48 Akrigg, D., Bleasby, A.J., Dix, N.I.M., Findlay, J.B.C., North, A.C.T., Parry-Smith, D., Wootton, J.C., Blundell, T.L., Gardner, S.P., Hayes, F., Islam, S., Sternberg, M.J.E., Thornton, J.M. and Tickle, I.J., *Nature*, 335 (1988) 745.
- 49 Huysmans, M., Richelle, J. and Wodak, S.J., *Proteins*, 11 (1991) 59.
- 50 Vriend, G., Stouten, P. and Sander, C., submitted.
- 51 Levitt, M., *J. Mol. Biol.*, 226 (1992) 507.
- 52 Vriend, G. and Sander, C., submitted.
- 53 Singh, J. and Thornton, J.M., *J. Mol. Biol.*, 211 (1990) 595.
- 54 Vriend, G. and Sander, C., *J. Appl. Cryst.*, 26 (1993) 47.
- 55 Pakula, A.A. and Sauer, R.T., *Annu. Rev. Genet.*, 23 (1989) 289.
- 56 Brennan, R.G., Wozniak, J., Faber, R. and Matthews, B.W., *J. Cryst. Growth*, 90 (1988) 160.
- 57 Van Gunsteren, W.F., *Protein Eng.*, 2 (1989) 5.
- 58 Yun-Yu, S., Mark, A.E., Cun-xin, W., Funua, H., Berendsen, H.J.C. and Van Gunsteren, W.F., *Protein Eng.*, 6 (1993) 289.
- 59 Priest, F.G., In Harwood, C.R. (Ed.), Biotechnology Handbook, Plenum Press, 1989, pp. 293–320.
- 60 Wetmore, D.R., Wong, S.L. and Roche, R.S., *Mol. Microbiol.*, 6 (1992) 1593.
- 61 Wandersman, C., *Mol. Microbiol.*, 3 (1989) 1825.
- 62 Heinrikson, R.L., *Methods Enzymol.*, 47 (1977) 175.
- 63 Stoeva, S. and Kleinschmidt, T., *Biol. Chem. Hoppe-Seyler*, 370 (1989) 1139.
- 64 Titani, K., Hermodson, M.A., Ericsson, L.H., Walsh, K.A. and Neurath, H., *Nature*, 238 (1972) 35.
- 65 Takagi, M., Imanaka, T. and Aiba, S., *J. Bacteriol.*, 163 (1985) 824.
- 66 Sidler, W., Niederer, E., Suter, F. and Zuber, H., *Biol. Chem. Hoppe-Seyler*, 367 (1986) 643.
- 67 Yang, M., Ferrari, E. and Henner, D.J., *J. Bacteriol.*, 160 (1984) 15.
- 68 Vasantha, N., Thompson, L.D., Rhodes, G., Banner, C., Nagle, J. and Fipula, D., *J. Bacteriol.*, 159 (1984) 811.
- 69 Kubo, M. and Imanaka, T., *J. Gen. Microbiol.*, 134 (1988) 1883.
- 70 Matthews, B.W., Jansonius, J.N., Colman, P.M., Schoenborn, B.P. and Dupourque, D., *Nature*, 238 (1972) 37.
- 71 Holmes, M.A. and Matthews, B.W., *J. Mol. Biol.*, 160 (1982) 623.
- 72 Paupit, R.A., Karlson, R., Picot, D., Jenkins, J.A., Niklaus-Reimer, A.S. and Jansonius, J.N., *J. Mol. Biol.*, 199 (1988) 525.
- 73 Stark, W., Paupit, R.A., Wilson, K.S. and Jansonius, J.N., *Eur. J. Biochem.*, 207 (1992) 781.
- 74 Kester, W.R. and Matthews, B.W., *J. Biol. Chem.*, 252 (1977) 7704.
- 75 Hangauer, D.G., Monzingo, A.F. and Matthews, B.W., *Biochemistry*, 23 (1984) 5730.
- 76 Tronrud, D.E., Monzingo, A.F. and Matthews, B.W., *Eur. J. Biochem.*, 157 (1986) 261.
- 77 Toma, S., Campagnoli, S., De Gregoris, E., Gioanna, R., Margarit, I., Zamai, M. and Grandi, G., *Protein Eng.*, 2 (1989) 359.
- 78 Dunn, B.M., In Beynon, R.J. and Bond, J.S. (Eds), Proteolytic Enzymes, a Practical Approach, IRL Press, 1989, pp.

- 57-82.
- 79 Matthews, B.W., *Acc. Chem. Res.*, 21 (1988) 333.
- 80 Voordouw, G., Milo, C. and Roche, R.S., *Biochemistry*, 15 (1976) 3716.
- 81 Roche, R.S. and Voordouw, G., *CRC Crit. Rev. Biochem.*, (1978) 1.
- 82 Rasnick, D. and Powers, J.C., *Biochemistry*, 17 (1978) 4363.
- 83 Tronrud, D.E., Holden, H.M. and Matthews, B.W., *Science*, 235 (1987) 571.
- 84 Colman, P.M., Jansonius, J.N. and Matthews, B.W., *J. Mol. Biol.*, 70 (1972) 701.
- 85 Dalzoppo, D., Vita, C. and Fontana, A., *J. Mol. Biol.*, 182 (1985) 331.
- 86 Imanaka, T., Shibasaki, M. and Takagi, M., *Nature*, 324 (1986) 695.
- 87 Eijsink, V.G.H., Vriend, G., Van den Burg, B., Venema, G. and Stulp, B.K., *Protein Eng.*, 4 (1990) 99.
- 88 Eijsink, V.G.H., Van der Zee, J.R., Van den Burg, B., Vriend, G. and Venema, G., *FEBS Lett.*, 282 (1991) 13.
- 89 Vriend, G., Berendsen, H.J.C., Van der Zee, J.R., Van den Burg, B., Eijsink, V.G.H., *Protein Eng.*, 4 (1991) 941.
- 90 Eijsink, V.G.H., Vriend, G., Van der Zee, J.R., Van den Burg, B. and Venema, G., *Biochem. J.*, 285 (1992) 625.
- 91 Eijsink, V.G.H., Vriend, G., Van den Burg, B., Van der Zee, J.R. and Venema, G., *Protein Eng.*, 5 (1992) 165.
- 92 Eijsink, V.G.H., Vriend, G., Van den Burg, B., Van der Zee, J.R., Veltman, O.R., Stulp, B.K. and Venema, G., *Protein Eng.*, 5 (1992) 157.
- 93 Toma, S., Campagnoli, S., Margarit, I., Gianna, R., Grandi, G., Bolognesi, M., De Filippis, V. and Fontana, A., *Biochemistry*, 30 (1991) 97.
- 94 Margarit, I., Campagnoli, S., Frigerio, F., Grandi, G., De Filippis, V. and Fontana, A., *Protein Eng.*, 5 (1992) 543.
- 95 Hardy, F., Vriend, G., Veltman, O.R., Van der Vinne, B., Venema, G. and Eijsink, V.G.H., *FEBS Lett.*, 317 (1993) 89.
- 96 Van den Burg, B., Dijkstra, B.W., Vriend, G., Van der Vinne, B., Venema, G. and Eijsink, V.G.H., submitted.
- 97 Privalov, P.L., *Adv. Prot. Chem.*, 33 (1979) 167.
- 98 Pace, C.N., *Trends Biochem. Sci.*, 15 (1990) 14.
- 99 Creighton, T.E., *Biochem. J.*, 270 (1990) 1.
- 100 Pace, C.N., *Methods Enzymol.*, 131 (1986) 266.
- 101 Pace, C.N., *Trends Biotech.*, 8 (1990) 93.
- 102 Becktel, W.J. and Baase, W.A., *Biopolymers*, 26 (1986) 619.
- 103 Becktel, W.J. and Schellman, J.A., *Biopolymers*, 26 (1987) 1859.
- 104 Segawa, S.I. and Sugihara, M., *Biopolymers*, 23 (1984) 2473.
- 105 Matouschek, A., Kellis, Jr., J.T., Serrano, L. and Fersht, A.R., *Nature*, 340 (1989) 122.
- 106 Chen, B., Baase, W.A. and Schellman, J.A., *Biochemistry*, 28 (1989) 691.
- 107 Sancho, J., Meiering, E.M. and Fersht, A.R., *J. Mol. Biol.*, 221 (1991) 1007.
- 108 Jackson, S.E. and Fersht, A.R., *Biochemistry*, 30 (1991) 10436.
- 109 Liang, H. and Terwilliger, T.C., *Biochemistry*, 30 (1991) 2772.
- 110 Serrano, L., Matouschek, A. and Fersht, A.R., *J. Mol. Biol.*, 224 (1992) 805.
- 111 Careaga, C.L. and Falke, J.J., *J. Mol. Biol.*, 226 (1992) 1219.
- 112 Dahlquist, F.W., Long, J.W. and Bigbee, W.L., *Biochemistry*, 15 (1976) 1103.
- 113 Wells, J.A. and Powers, D.B., *J. Biol. Chem.*, 261 (1986) 6564.
- 114 Van den Burg, B., Eijsink, V.G.H., Stulp, B.K. and Venema, G., *Biochem. J.*, 272 (1990) 93.
- 115 Eijsink, V.G.H., Van den Burg, B., Vriend, G., Berendsen, H.J.C. and Venema, G., *Biochem. Int.*, 24 (1991) 517.
- 116 Wetzel, R., *Trends Biochem. Sci.*, 12 (1987) 478.
- 117 Mitchinson, C. and Wells, J.A., *Biochemistry*, 28 (1989) 480.
- 118 McLendon, G. and Radany, E., *J. Biol. Chem.*, 252 (1978) 6335.
- 119 Pace, C.N. and Barrett, A.J., *Biochem. J.*, 219 (1984) 411.
- 120 Imoto, T., Yamada, H. and Ueda, T., *J. Mol. Biol.*, 190 (1986) 647.
- 121 Voordouw, G. and Roche, R.S., *Biochemistry*, 14 (1975) 4667.
- 122 Fassina, G., Vita, C., Dalzoppo, D., Zamai, M., Zambonin, M. and Fontana, A., *Eur. J. Biochem.*, 156 (1986) 221.
- 123 Van den Burg, B., Eijsink, V.G.H., Stulp, B.K. and Venema, G., *J. Biochem. Biophys. Methods*, 18 (1989) 209.
- 124 Daniel, R.M., Cowan, D.A., Morgan, H.W. and Curran, M.P., *Biochem. J.*, 207 (1982) 641.
- 125 Ogasahara, K., Tsunasawa, S., Soda, Y., Yutani, K. and Sugino, Y., *Eur. J. Biochem.*, 150 (1985) 17.
- 126 Parsell, D.A. and Sauer, R.T., *J. Biol. Chem.*, 264 (1989) 7590.

- 127 Fontana, A., Fassina, G., Vita, C., Dalzoppo, D., Zamai, M. and Zambonin, M., *Biochemistry*, 25 (1986) 1847.
- 128 Van Gunsteren, W.F. and Berendsen, H.J.C., BIOMOS, Biomolecular software, Laboratory of Physical Chemistry, University of Groningen, The Netherlands, 1987.
- 129 Hol, W.G.J. and Vriend, G., unpublished results.
- 130 Eijsink, V.G.H., Ph.D. Thesis, University of Groningen, The Netherlands, 1991.
- 131 Signor, G., Vita, C., Fontana, A., Frigerio, R., Bolognesi, M., Toma, S., Gianni, R., De Gregoris, E. and Grandi, G., *Eur. J. Biochem.*, 189 (1990) 221.
- 132 Frigerio, F., Vriend, G. and Eijsink, V.G.H., in preparation.
- 133 Alber, T., *Annu. Rev. Biochem.*, 58 (1989) 765.
- 134 Fontana, A., *Curr. Opin. Biotech.*, 2 (1991) 551.
- 135 Matthews, B.W., *Curr. Opin. Struct. Biol.*, 1 (1991) 17.
- 136 Sauer, R.T. and Lim, W., *Opin. Struct. Biol.*, 2 (1992) 46.
- 137 Argos, P., Rossmann, M.G., Grau, U.M., Zuber, H., Frank, G. and Tratschin, J.D., *Biochemistry*, 18 (1979) 5698.
- 138 Menedez-Arias, L. and Argos, P., *J. Mol. Biol.*, 206 (1989) 397.
- 139 Querrol, E. and Parilla, A., *Enzyme Microb. Technol.*, 9 (1987) 238.
- 140 Alber, T., Bell, J.A., Dao-Pin, S., Nicholson, H., Wozniak, J.A. and Matthews, B.W., *Science*, 239 (1988) 631.
- 141 Shortle, D., Sites, W.E. and Meeker, A.K., *Biochemistry*, 29 (1990) 8033.
- 142 Sandberg, W.S. and Terwilliger, T.C., *Proc. Natl. Acad. Sci. U.S.A.*, 88 (1991) 1706.
- 143 Serrano, L., Neira, J.L., Sancho, J. and Fersht, A.R., *Nature*, 356 (1992) 453.
- 144 Mrabet, N.T., Van den Broeck, A., Van den Brande, I., Stanssens, P., Laroche, Y., Lambeir, A.M., Matthijssens, G., Jenkins, J., Chiadmi, M., Van Tilbeurgh, H., Rey, F., Janin, J., Quax, W.J., Lasters, I., De Maeyer, M. and Wodak, S.J., *Biochemistry*, 31 (1992) 2239.
- 145 Takagi, H., Takahasi, T., Momose, H., Inouye, M., Maeda, Y., Matsuzawa, H. and Ohta, T., *J. Biol. Chem.*, 265 (1990) 6874.
- 146 Van der Vinne, B., Vriend, G., Venema, G. and Eijsink, V.G.H., in preparation.
- 147 Kauzmann, W., *Adv. Prot. Chem.*, 14 (1957) 1.
- 148 Tanford, C., *Adv. Prot. Chem.*, 23 (1968) 121.
- 149 Dill, K.A., *Biochemistry*, 29 (1990) 7133.
- 150 Sharp, K.A., *Curr. Opin. Struct. Biol.*, 1 (1991) 171.
- 151 Connolly, M.L., *J. Pept. Prot. Res.*, 28 (1986) 360.
- 152 Karpusas, M., Baase, W.A., Matsumura, M. and Matthews, B.W., *Proc. Natl. Acad. Sci. U.S.A.*, 86 (1989) 8237.
- 153 Lim, W.A. and Sauer, R.T., *Nature*, 339 (1989) 31.
- 154 Sandberg, W.C. and Terwilliger, T.C., *Science*, 245 (1989) 54.
- 155 Rashin, A.A., Iofin, M. and Honig, B., *Biochemistry*, 25 (1986) 3619.
- 156 Creighton, T.E., *Curr. Opin. Struct. Biol.*, 1 (1991) 5.
- 157 Kellis, Jr., J.T., Nyberg, K., Sali, D. and Fersht, A.R., *Nature*, 333 (1988) 784.
- 158 Eriksson, A.E., Baase, W.A., Zhang, X.J., Heinz, D.W., Blaber, M., Baldwin, E.P. and Matthews, B.W., *Science*, 255 (1992) 178.
- 159 Matsumura, M., Becktel, W.J. and Matthews, B.W., *Nature*, 334 (1988) 406.
- 160 Kimura, S., Oda, Y., Nakai, T., Katayanagi, K., Kitakuni, E., Nakai, C., Nakamura, H., Ikehara, M. and Kanaya, S., *Eur. J. Biochem.*, 206 (1992) 337.
- 161 Voorintholt, R., Kosters, M.T., Vegter, G., Vriend, G. and Hol, W.G.J., *J. Mol. Graphics*, 7 (1989) 243.
- 162 Richardson, J.S. and Richardson, D.C., *Science*, 240 (1988) 1648.
- 163 Presta, L.G. and Rose, G.D., *Science*, 240 (1988) 1632.
- 164 Wada, A., *Adv. and Biophys.*, 9 (1976) 1.
- 165 Hol, W.G.J., Van Duijnen, P.T. and Berendsen, H.J.C., *Nature*, 273 (1978) 443.
- 166 Aquist, J., Luecke, H., Quiocco, F. and Warshel, A., *Proc. Natl. Acad. Sci. U.S.A.*, 88 (1991) 2026.
- 167 Sali, D., Bycroft, M. and Fersht, A.R., *Nature*, 335 (1988) 740.
- 168 Nicholson, H., Becktel, W.J. and Matthews, B.W., *Nature*, 336 (1988) 651.
- 169 Serrano, L. and Fersht, A.R., *Nature*, 342 (1989) 296.
- 170 Dao-Pin, S., Sauer, U., Nicholson, H. and Matthews, B.W., *Biochemistry*, 30 (1991) 7142.
- 171 Sali, D., Bycroft, M. and Fersht, A.R., *J. Mol. Biol.*, 220 (1991) 779.

- 172 Sancho, J., Serrano, L. and Fersht, A.R., *Biochemistry*, 31 (1992) 2253.
- 173 Bell, J.A., Becktel, W.J., Sauer, U., Baase, W.A. and Matthews, B.W., *Biochemistry*, 31 (1992) 3590.
- 174 Perutz, M.F. and Raidt, H., *Nature*, 255 (1975) 256.
- 175 Marqusee, S. and Baldwin, R.L., *Proc. Natl. Acad. Sci. U.S.A.*, 84 (1987) 8898.
- 176 Bradley, E.K., Thomason, J.F., Cohen, F.E., Kosen, P. and Kuntz, I.D., *J. Mol. Biol.*, 215 (1990) 607.
- 177 Lyu, P.C., Gans, P.J. and Kallenbach, N.R., *J. Mol. Biol.*, 223 (1992) 343.
- 178 Akke, M. and Forssen, S., *Proteins*, 8 (1990) 23.
- 179 Horovitz, A., Serrano, L., Avron, B., Bycroft, M. and Fersht, A.R., *J. Mol. Biol.*, 216 (1990) 1031.
- 180 Serrano, L., Horovitz, A., Avron, B., Bycroft, M. and Fersht, A.R., *Biochemistry*, 29 (1990) 9343.
- 181 Anderson, D.E., Becktel, W.J. and Dahlquist, F.W., *Biochemistry*, 29 (1990) 2403.
- 182 Erwin, C.R., Barnett, B.L., Oliver, J.D. and Sullivan, J.F., *Protein Eng.*, 4 (1990) 87.
- 183 Goldenberg, D.P. and Creighton, T.E., *Biopolymers*, 24 (1985) 167.
- 184 Alber, T., Dao-Pin, S., Wilson, K., Wozniak, J.A., Cook, S.P. and Matthews, B.W., *Nature*, 330 (1987) 41.
- 185 Perry, K.M., Onuuffer, J.J., Touchette, N.A., Herndon, C.S., Gittelman, M.S., Matthews, C.R., Chen, J.T., Mayer, R.J., Taira, K., Benkovic, S.J., Howell, E.E. and Kraut, J., *Biochemistry*, 26 (1987) 2674.
- 186 Bryan, P.N., Rollence, M.L., Pantoliano, M.W., Wood, J., Finzel, B.C., Gilliland, G.L., Howard, A.J. and Poulos, T.L., *Proteins*, 1 (1986) 326.
- 187 Pantoliano, M.W., Whitlow, M., Wood, J.F., Dodd, S.W., Hardman, K.D., Rollence, M.L. and Bryan, P.N., *Biochemistry*, 28 (1989) 7205.
- 188 Vriend, G., *Protein Eng.*, 4 (1990) 221.
- 189 Matthews, B.W., Nicholson, H. and Becktel, W.J., *Proc. Natl. Acad. Sci. U.S.A.*, 84 (1987) 6663.
- 190 Pantoliano, M.W., Whitlow, M., Wood, J.F., Rollence, M.L., Finzel, B.C., Gilliland, G.L., Poulos, T.L. and Bryan, P.N., *Biochemistry*, 27 (1988) 8311.
- 191 Kuroki, R., Kawakita, S., Nakamura, H. and Yutani, K., *Proc. Natl. Acad. Sci. U.S.A.*, 89 (1992) 6803.
- 192 Hecht, M.H., Sturtevant, J.M. and Sauer, R.T., *Proteins*, 1 (1986) 43.
- 193 Ganter, C. and Pluckthun, A., *Biochemistry*, 29 (1990) 9395.
- 194 Finkelstein, A.V., Badretdinov, A.Y. and Ptitsyn, O.B., *Proteins*, 10 (1991) 287.
- 195 O'Neill, K.T. and DeGrado, W.F., *Science*, 250 (1990) 646.
- 196 Dao-Pin, S., Baase, W.A. and Matthews, B.W., *Proteins*, 7 (1990) 198.
- 197 Zhang, X.J., Baase, W.A. and Matthews, B.W., *Biochemistry*, 30 (1991) 2012.
- 198 Suzuki, Y., Oishi, K., Nakano, H. and Nagayama, T., *Appl. Microbiol. Biotechnol.*, 26 (1987) 546.
- 199 Tesfay, H.S., Amelunxen, R.E. and Goldberg, I.D., *Gene*, 82 (1989) 237.
- 200 Yutani, K., Hayashi, S., Sugisaki, Y. and Ogasahara, K., *Proteins*, 9 (1991) 90.
- 201 Van Osten, K., personal communication.
- 202 Pace, C.N., Grimsley, G.R., Thomson, J.A. and Barnett, B.J., *J. Biol. Chem.*, 263 (1988) 11820.
- 203 Jacobson, R.H., Matsumura, M., Faber, H.R. and Matthews, B.W., *Protein Sci.*, 1 (1992) 46.
- 204 Matsumura, M., Becktel, W.J., Levitt, M. and Matthews, B.W., *Proc. Natl. Acad. Sci. U.S.A.*, 86 (1989) 6562.
- 205 Kanaya, S., Katsuda, C., Kimura, S., Nakai, T., Kitakuni, E., Nakamura, H., Katayanagi, K., Morikawa, K. and Ikebara, M., *J. Biol. Chem.*, 266 (1991) 6038.
- 206 Gusev, N.B., Grabarek, Z. and Gergely, J., *J. Biol. Chem.*, 266 (1991) 16622.
- 207 Matsumura, M., Signor, G. and Matthews, B.M., *Nature*, 342 (1989) 291.
- 208 Pantoliano, M.W., Ladner, R.C., Bryan, P.N., Rollence, M.L., Wood, J.F. and Poulos, T.L., *Biochemistry*, 26 (1987) 2077.
- 209 Van den Burg, B., Dijkstra, B.W., Van der Vinne, B., Stulp, B.K., Eijlsink, V.G.H. and Venema, G., *Protein Eng.*, 6 (1993) in press.
- 210 Wells, J.A., *Biochemistry*, 29 (1990) 8509.
- 211 Pfeill, W., *Biophys. Chem.*, 13 (1981) 181.
- 212 Corbett, R.J.T., Ahmad, F. and Roche, R.S., *Biochem. Cell Biol.*, 64 (1986) 953.
- 213 Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, Jr., M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasuma, M., *J. Mol. Biol.*, 112 (1977) 535.
- 214 Takagi, M. and Imanaka, T., *FEBS Lett.*, 254 (1989) 43.
- 215 Carson, M., *J. Mol. Graphics*, 5 (1987) 103.