Request ID: DDS36734
User: Gangi-Dino, Rita
Location: ILL
Requested on: 11/25/2005
Needed by: 11/30/2005
Journal Title: Proteins
ISSN: 0887-3585
Article Author(s): Frommel C
Article Title: Thermitase, a thermostable subtilisin: comparison of predicted and experimental structures and the m
Year: 1989
Volume: 5
Issue: 1
Pages: 22-37
PMID: 2664764
User's Comments: in color, if avilable
Staff Notes: CI-5880022

CI-05880022-9

DD704731*

DD704731

CISTI ICIST

CI-05880022-9

Document Delivery Service
in partnership with the **Canadian Agriculture Library**

Service de fourniture de Documents
en collaboration avec la **Bibliothèque canadienne de l'agriculture**

THIS IS NOT AN INVOICE / CECI N'EST PAS UNE FACTURE

ANTHONY ARTALE
MED LIB NATHAN CUMMINGS CTR (S-46)
MEMORIAL SLOAN KETTERING CANCER CTR
1275 YORK AVENUE
NEW YORK, NY 10021
UNITED STATES

ORDER NUMBER:	CI-05880022-9
Account Number:	DD704731
Delivery Mode:	ARI
Delivery Address:	arielsf.infotrieve.com/140.163.217.217
Submitted:	2005/11/25 11:08:20
Received:	2005/11/25 11:08:20
Printed:	2005/11/25 11:52:42

Direct	**Periodical**	**OPENURLOPAC**	**UNITED STATES**

Client Number:	DDS36734/GANGI-DINO, RITA
Title:	**PROTEINS.**
DB Ref. No.:	IRN10402846
ISSN:	ISSN08873585
Vol./Issue:	5/1
Date:	1989
Pages:	22-37
Article Title:	THERMITASE, A THERMOSTABLE SUBTILISIN
Article Author:	FROMMEL C
Report Number:	IRN10402846
Publisher:	ALAN R. LISS,

Estimated cost for this 16 page document: $10.2 document supply fee + $30 copyright = $40.2

Phone/Téléphone: 1-800-668-1222 (Canada - U.S./E.-U.) (613) 998-8544 (International)
www.nrc.ca/cisti Fax/Télécopieur: (613) 993-7619 www.cnrc.ca/icist
info.cisti@nrc.ca info.icist@nrc.ca

National Research Council Canada Conseil national de recherches Canada

PROTEINS: Structure, Function, and Genetics 5:22–37(1989)

Thermitase, a Thermostable Subtilisin: Comparison of Predicted and Experimental Structures and the Molecular Cause of Thermostability

Cornelius Frömmel and Chris Sander
Biocomputing Programme, EMBL, D-6900 Heidelberg, Federal Republic of Germany

ABSTRACT **The subtilisin family of proteases has four members of known sequence and structure: subtilisin Carlsberg, subtilisin novo, proteinase K, and thermitase. Using thermitase as a test case, we ask two questions. How good are methods for model building a three-dimensional structure of a protein based on sequence homology to a known structure? And what are the molecular causes of thermostability? First, we compare predicted models of thermitase, refined by energy minimization and varied by molecular dynamics, with the preliminary crystal structure. The predictions work best in the conserved structural core and less well in seven loop regions involving insertions and deletions relative to subtilisin. Here, variation of loop regions by molecular dynamics simulation *in vacuo* followed by energy minimization does not improve the prediction since we find no correlation between *in vacuo* energy and correctness of structure when comparing local energy minima. Second, in order to identify the molecular cause of thermostability we confront hypotheses derived by calculation of the details of interatomic interactions and estimates of hydrophobic interactions with inactivation experiments. As a result, we can exclude salt bridges and hydrophobic interactions as main causes of thermostability. Based on a combination of theoretical and experimental evidence, the unusually tight binding of calcium by thermitase emerges as the most likely single influence responsible for its increased thermostability.**

Key words: sequence homology, tertiary structure prediction, molecular dynamics, energy minimization, hydrophobic interactions, aromatic ring–ring interactions, salt bridges, calcium binding, thermoactinomyces vulgaris, extracellular protease

INTRODUCTION

Subtilisin has for decades been one of a handful of technologically mass-produced enzymes, with world production of several hundred tons in 1983.[1] The enzyme loses its function by irreversible inactivation with a half-life in the order of weeks at 20°C and in the order of minutes at 60°C (pH 7–9).[2] There is considerable practical interest in increasing its stability. One group has replaced Met-222 by 19 other amino acids: while oxidation was prevented, enzymatic activity decreased significantly.[3] Others have mutated Ser-24 and Ser-87 to produce a stabilizing disulfide link, with controversial results: one group did not achieve stabilization with this mutation[4] while another did.[5] In other experiments, a mutation of Asn-218 in SBT to Ser-218 resulted in an enzyme with 4-fold higher stability.[6] However, a systematic theoretical guide to stability-enhancing point mutations is not yet available.

Until the advent of a general theory of thermostability much can be learned from natural thermostable variants. In the subtilisin family there are two members that are more stable than subtilisin novo (SBT) and subtilisin Carlsberg (SCB): proteinase K (PRK) and thermitase (TRM), with TRM the most stable. Three enzymes in the family (SCB, SBT, and PRK) have been known in sequence and three-dimensional structure for some time. Thermitase itself was first isolated in 1978,[7] its amino acid sequence determined in 1984,[8] the enzyme crystallized for X-ray analysis in 1985, and the polypeptide chain traced in an electron density map based on data up to 2.5 Å resolution.[9] Experiments in a number of environmental conditions have illustrated its unusual stability.[10] The stability varies appreciably with the concentration of calcium ions in the range 10^{-5} to 10^{-3}, but until now it has been impossible to remove one strongly bound calcium ion (L. Briedigkeit and C. Frömmel, unpublished results) without denaturation, despite numerous attempts by methods usually successful with other proteins that strongly bind calcium.[11]

Here, we carefully check several hypotheses about

Received March 1, 1988; accepted December 20, 1988.

Address reprint requests to Chris Sander, Biocomputing Programme, EMBL, Meyerhofstrasse, 1, D-6900 Heidelberg, Federal Republic of Germany.

Cornelius Frömmel's permanent address is Institute of Biochemistry, Humboldt-University, Hessische Strasse 3-4, 1040 Berlin, German Democratic Republic.

which factors contribute most to the stability of thermitase, in part by theoretical calculation and in part by experiment, and we assess the predictability of such factors through model building by homology.

MATERIALS AND METHODS

Alignment of Sequences and Structures

Our computer program for alignment of sequences (Table I), MaxHomology (C. Sander, 1982, unpublished), is based on the algorithm of Smith and Waterman[12] and uses the similarity matrix of McLachlan.[13] For the pairwise sequence alignments in Table I the gap penalty was 2.00 single residue matches for the first residue in a gap, and 0.1 for each additional residue in a gap.

Independently of the sequence alignment, the three-dimensional structures of SBT, SCB, and PRK were aligned by identifying conserved substructures. Step 1: Find locally optimal superpositions of main chain atoms in fixed-length fragments of protein A in protein B. Step 2: Retain fragment pairs only if their superposition is extendable to the nonloop parts of the entire protein (helices and strands) to within 2.0 Å rms deviation of C(α) positions; if there are several alternatives for a fragment pair, retain the one with the lowest rms. The overlaps of retained fragments in all three proteases are called structurally conserved segments; their union is called the conserved core.

Protein structures were viewed and compared interactively at the Max-Planck-Institut für Medizinishce Forschung in Heidelberg, FRG, using Max-Grafix, version CELLO-86 (C. Sander, unpublished), on the vector calligraphic display WIRRWARR (W. Gebhard and W. Kabsch, unpublished). Coordinates of SBT (BPN') were taken as data set 1SBT from the Protein Data Bank, Brookhaven, N.Y., U.S.A.[14]; C(alpha) coordinates of PRK[15] are a gift from Wolftram Sänger, FU Berlin, FRG; preliminary all-atom coordinates SCBØ (Carlsberg) and SBTØ (subtilisin BPN') are a prepublication gift of Mike James and Catherine A. McPhalen, Dept. of Biochemistry, University of Alberta, Edmonton, Canada. The preliminary coordinates of thermitase are from the Institute of Crystallography, USSR Academy of Science, Moscow (Teplyakov et al. [9] and A. W. Tepliakov, personal communication).

Model Building by Homology

The model of the three-dimensional structure of TRM was built well before the experimental structure was available following these steps[16]: replacement of homologous residues, deletions and insertions, regularization, steric relaxation (Fig. 1A and B). The starting point was the coordinates of subtilisin novo (1SBT). Starting conformations for loops were model built intuitively with steric criteria and standard loop conformations in mind, without attempting to evaluate all combinatorially possible arrangements (details in Table I and Fig. 2). The exchange of Val-8 (1SBT) by Pro-15 (TRM) was possible without changing main chain angles.

Refinement of Model and Generation of Alternate Loop Conformations

The rough coordinates (TRMØ, Fig. 1B) were the starting point of geometric and energy regularization (result: TRM1). In order to generate alternate models, up to 17.5 psec of molecular dynamics trajectories was generated at 300° Kelvin, starting from the energy-minimized first model (TRM1, Fig. 4A and B), using van Gunsteren and Berendsen's software, GROMOS 86[17] (update of neighbors within a cutoff radius of 8 Å every 10 cycles; time step, 2 fsec; positional harmonic constraints, when used, with a force constant of 10 kcal/mol Å^2). As a control, we applied these methods to the known SBTØ structure as well.

Estimation of Energy Differences

Differences in intramolecular potential energy were calculated theoretically and differences in the free energy of stabilization estimated on the basis of inactivation experiments. Intramolecular energy terms in the GROMOS potentials as used here are electrostatic potential with partial charges on polar atoms, 6–12 nonbonded potential for all atoms and terms associated with bond lengths, bond angles, and torsion angles. Effects not handled properly in energy calculations *in vacuo* using GROMOS include salt bridges, special aspects of ring–ring interactions, and protein–water interactions.

We explicitly estimated these additional terms. Ionic charge–charge interactions were counted in an all-or-none fashion, depending on whether the interacting groups of atoms are within a cutoff distance or not. Ring–ring interactions were treated similarly to account for possible ring–ring stabilization beyond the simple atom–atom dispersion energy term.[18,19] Both charge–charge and ring–ring interactions are reported as distance plots (Fig. 5). Protein–water interaction was not estimated as a separate term that can be added to the nonbonded energy calculated within GROMOS. Instead, we independently estimated the hydrophobic contribution to the free energy difference of folding in an aqueous medium by exploiting an empirical relationship between apolar surface area and the transfer energy of individual amino acid side chains between water and octanol, as calibrated recently by Frömmel[20]: 0.017 kcal/mol per Å^2 of apolar surface. The unfolded reference state for amino acid X is taken from the extended conformation of the oligopeptide G-G-X-G-G. The free energy difference of hydrophobic stabilization estimated in this way and the enthalpic nonbonded energy difference derived from GROMOS are not simply additive since the free energy transfer to octanol also contains an enthalpic term.

Protein Purification

Thermitase (TRM, EC 3.4.21.14), an extracellular protease from *Thermoactinomyces vulgaris*, was purified from concentrated culture filtrate by flat bed isoelectric focusing.[7] Subtilisin Carlsberg (SCB) was purchased from sigma, incorrectly labeled subtilisin BPN' (P 5255), but recently identified as subtilisin Carlsberg from *Bacillus licheniformis*.[21] The identity of all preparations of subtilisins used in experiments was confirmed by CNBr cleavage followed by molecular weight analysis of the peptides using sodium dodecyl sulfate electrophoresis.[22] Absolute protein concentration was determined by a modified biuret method[23], using bovine serum albumin as a standard, or spectrophotometrically at 280 nm using absorption coefficient A[1%, 1 cm] with $A = 18.5$ for TRM,[7] and $A = 11.7$ for SCB.[24] All other chemical reagents are from Laborchemie, Apolda, GDR, and were at least analytical grade. The residual calcium content of thermitase was at least 0.8 ions per molecule while SCB was calcium free after 30 minute incubation on chelex resin or after 2 hour dialysis against 10 mM EDTA (L. Briedigkeit, C. Frömmel, unpublished).

Activity Measurements

Esterolytic activity of thermitase was determined photometrically at 400 nm wavelength using *p*-nitrophenyl acetate (pNPA) as substrate (0.5 mM pNPA in 0.1 M Tris–HCl, at pH 8.0, with 1% acetone and 20 µl enzyme solution at a concentration of about 1 mg/ml, at 25°C). Proteolytic activity was monitored with casein as substrate (Hammersten, E. Merck, Darmstadt, FRG) as in Frömmel et al.[7]

Inactivation Measurements

Loss of function of the subtilisin proteases is due to denaturation and/or autolysis. Autolysis requries unfolding as the initial step.[10,25] Both types of inactivation are completely irreversible. The time course of inactivation was determined by incubating the enzyme (about 1 mg/ml enzyme in either boronic acid/NaOH buffer pH 9.0, or, in acetic acid/NaOH buffer at pH 3.5) at 60°C and measuring the residual enzyme activity at 25°C. Rate constants for inactivation were then extracted from the first-order (lower pH value, denaturation is the rate-limiting process) and second-order (higher value, autolysis is the rate-limiting step) plot, respectively. In both types of experiment only small peptide fragments remain in solution after inactivation.[10] The changes in the rate of inactivation at different concentrations of ligands and cosolutes was converted to free energy of stabilization[26] (Fig. 6):

$$G(\text{stab}) = RT \ln [k(0)/k]$$

where k is the inactivation rate constant at a given ligand or cosolute concentration and $k(0)$ is the inactivation rate constant measured under reference conditions. ΔG_{stab} represents the change in activation energy due to the added substances, where the activation energy is the difference between the free energy of the ground state and the activated state.

RESULTS AND DISCUSSION

We present three groups of results: (1) a direct comparison of TRM model structures and its experimentally determined conformation in order to assess the reliability of model building; (2) hypotheses based on the comparison of TRM with its brother enzymes, SCB, SBT, or PRK, as to which molecular interactions determine structural stability (e.g., salt bridges, ring–ring interactions, hydrophobic interactions); (3) evaluation of the hypotheses by confrontation with experimental results. Finally, we discuss general theories of thermostability.

Overall Accuracy of TRM Models

With current techniques, how well can one expect to be able to predict a three-dimensional structure built on the basis of sequence homology to known structures? Due to the conservation of protein cores during evolution there is little doubt that a model based on 40% sequence homology has essentially correct assignments of large parts of the protein and an essentially correct core structure (see Table I and Fig. 1A). Quantitatively, from the analysis of conserved cores of protein pairs by Lesk and Chothia,[28,29] the 40–50% homology between TRM and SBT/SCB/PRK (Table I) indicates that the correct structure of TRM differs from the known structures of SBT/SCB/PRK and from any rough model based on these by about 0.6–1.6 Å rms difference of C(α) atom positions in the core. Comparison of the model structure with the unrefined X-ray structure confirmed these expectations. The actually observed C(α) rms deviation between model (TRM1) and experimental structure, after optimal superposition (Fig. 3A), is 1.2 Å in the core.

Improvement by Molecular Dynamics?

Can the starting model be improved by molecular dynamics? The 17.5 psec constrained molecular dynamics trajectory, starting with the first model, does provide a set of alternative model conformations (Fig. 4A and B). These conformations scatter about that of the first model with an rms of typically 2.3 Å for C(α) atoms and 3.5 Å for all atoms. In loops, C(α) atoms move away from the starting position up to 5 Å (restraints on core) and 6 Å (unrestrained calculations)

Is nonbonded energy a useful criterion for the selection of correct conformations from among these alternative structures? Energy minimization (*in vacuo*) of the starting structure, of several snapshots (10 per psec) during the trajectory and of conformations time averaged over several psec results in non-

TABLE IA. Sequence Alignment of Four Subtilisin Proteases and Segments Conserved in the Three-Dimensional Structure

```
                                                    II
TRM YTPNDPYFSSR-QYGPQKIQAPQAW-DIAE----- GSGAKIAIVDTGVQSNHPDL AGKVVG    1- 54
SBT -------AQSV-PYGVSQIKAPALHSQGYT-----<GSNVKVAVIDSGIDSSHPDL>--KVAG    1- 46
PRK ------AAQTNAPWGLARISSTSPGTSTYYYDESA<GQGSCVYVIDTGIEASHPEF>EGRAQ-    1- 55
SCB -------AQTV-PYGIPLIKADKVQAQGFK-----<GANVKVAVLDTGIQASHPDL>--NVVG    1- 46
          3322 33*   *232     222      <*22233332*3*3223**33> 22323

                          III                          IV                V
TRM GWDFVDNDSTP-Q NGNGHGTHCAGIAAA VTNNSTGIA GTAPKASILAVRVL DNSGSG TWT 55-113
SBT GASMVPSETNPFQ<DDNSHGTHVAGTVAA>L-NNSIGVL<GVAPSASLYAVKVL>GADGSG<QYS 47-105
PRK ---MVKTYYYSSR<DGNGHGTHCAGTVGS>---GRTY--<GVAKKTQLFGVKVL>DDNGSG<QYS 56-106
SCB GASFVAGEAYN-T<DGNGHGTHVAGTVAA>L-DNTTGVL<GVAPSVSLYAVKVL>NSSGSG<SYS 47-104
    3222*     22 2<33*3****2***333>2 2323322<*3*3223322*3**>2 2***<233

      V                    VI                        VII
TRM AVANGITYAADQD ---A---- KVISLSLGG TVGNS GLQQAVNYAWNKGSVVVAAAGNA   114-164
SBT WIINGIEWAIANN>---M-----<DVINMSLGG>PSGSA<ALKAAVDKAVASGVVVVAAAGNE>  106-156
PRK TIIAGMDFVASDK>NNRNCPKG<VVASLSLGG>GY-SS<SVNSAAARLQSSGVMVAVAAGNN>   107-163
SCB GIVSGIQWATTNG>---M-----<DVINMSLGG>ASGST<AMKQAVDNAYARGVVVVAAAGNS>  105-155
     322*3 232 2 >    2     <2*322****>  332<2222*32 3 22*33*33**** >

                          VIII
TRM GNTAP----NY PAYYSNAIAVASTDQNDNKSSFSTYGSVVDVAAPG SW LYSTYPTS       165-215
SBT GSTGSSSTVGY<PGKYPSVIAVGAVDSSNQRASFSSVGPELDVMAPG>VS<IQSTLPGN       157-212
PRK NA--DAR-NYS<PASEPSVCTVGASDRYDRRSSFSNYGSVLDIFGPG>--<TDSLWIGG       164-213
SCB GNSGSTNTIGY<PAKYDSVIAVGAVDSNSNRASFSSVGAELEVMAPG>AG<VYSTYPTN       156-211
    32222  2 23<*32323333**32*222232***32*2233323**>  <22*32322

                  IX
TRM TYASLSGTSMATPHVAGVAGLLASQG --RSASNIRAAIENTADKISGTGTYWAKGRV         217-271
SBT KYGAYNGTSMASPHVAGAAALILSKH>PNWTNTQVRSSLENTTTKLGDSFYY-GKGLI        213-268
PRK STRSISGTSMATPHVAGLAAYLMTLG>KTTAASACRY-IADTANK-GDLSNI-PFGTV        214-267
SCB TYATLNGTSMASPHVAGAAALILSKH>PNLSASQVRNRLSSTATYLGSSFYY-GKGLI        212-267
    232222*****2*****2**322322>22 23322*  222*3232322223 22*22

                                                                       272-279
TRM NAYKAVQY                                                           269-275
SBT NVQAAAQ                                                            268-276
PRK NLAYNNYQA                                                          268-274
SCB NVQAAAQ                                                            
    *222323
```

*: Residues conserved identically in all four sequences; 3: Residues conserved identically in three of the four sequences; 2: Residues conserved identically in two of the four sequences; < >: Conserved segments in the three-dimensional structures of SBT, PRK, and SCB. Taken together, these segments are the "conserved core" (see Fig. 1A, B, and Fig. 2). References for sequence data: SBT,[53,54] SCB,[55] TRM,[8] PRK.[56]

TABLE IB. Sequence Similarity (% Identity) Between the Four Subtilisin Proteases*

	SBT	SCB	TRM	PRK
SBT		70	43	37
SCB	78		47	39
TRM	52	55		37
PRK	48	48	44	

*Top right, all residues; bottom left, residues in conserved core only.

bonded locally minimal energies (GROMOS) of −3.3 to -3.5×10^3 kcal/mol; these energies are comparable to those from the energy-minimized high-resolution structures of SBTØ and SCBØ. Molecular dynamics does improve the model structure in that the topology of loops generated during the dynamics trajectory are at times closer to the experimental structure than the starting conformation (one example is in Table II). However, since we have no correlation between lower rms deviation from the correct structure and lower nonbonded *in vacuo* energy, we have no obvious criterion for picking out these transiently correct conformations.

Our technique for exploration of alternate loop conformations by trajectory dynamics is similar to that of Fine et al.[30] (although less general, as we only used one starting conformation rather than randomly generated ones) and simpler than the combinatorially exhaustive of Moult and James[31] that in its current form becomes prohibitive for loops longer than four to five residues. We confirm[30] one problem of the dynamics simulations *in vacuo*, an unrealistic decrease of electrostatic potential energy: in unconstrained runs the electrostatic energy deceased by 5–25% relative to zero (less in constrained runs). While the local interaction parame-

Abbreviations: TRM, thermitase; SCB, subtilisin Carlsberg; SBT, subtilisin BPN′ alias novo; PRK, proteinase K; EDTA, ethylenediaminetetraacetic acid.

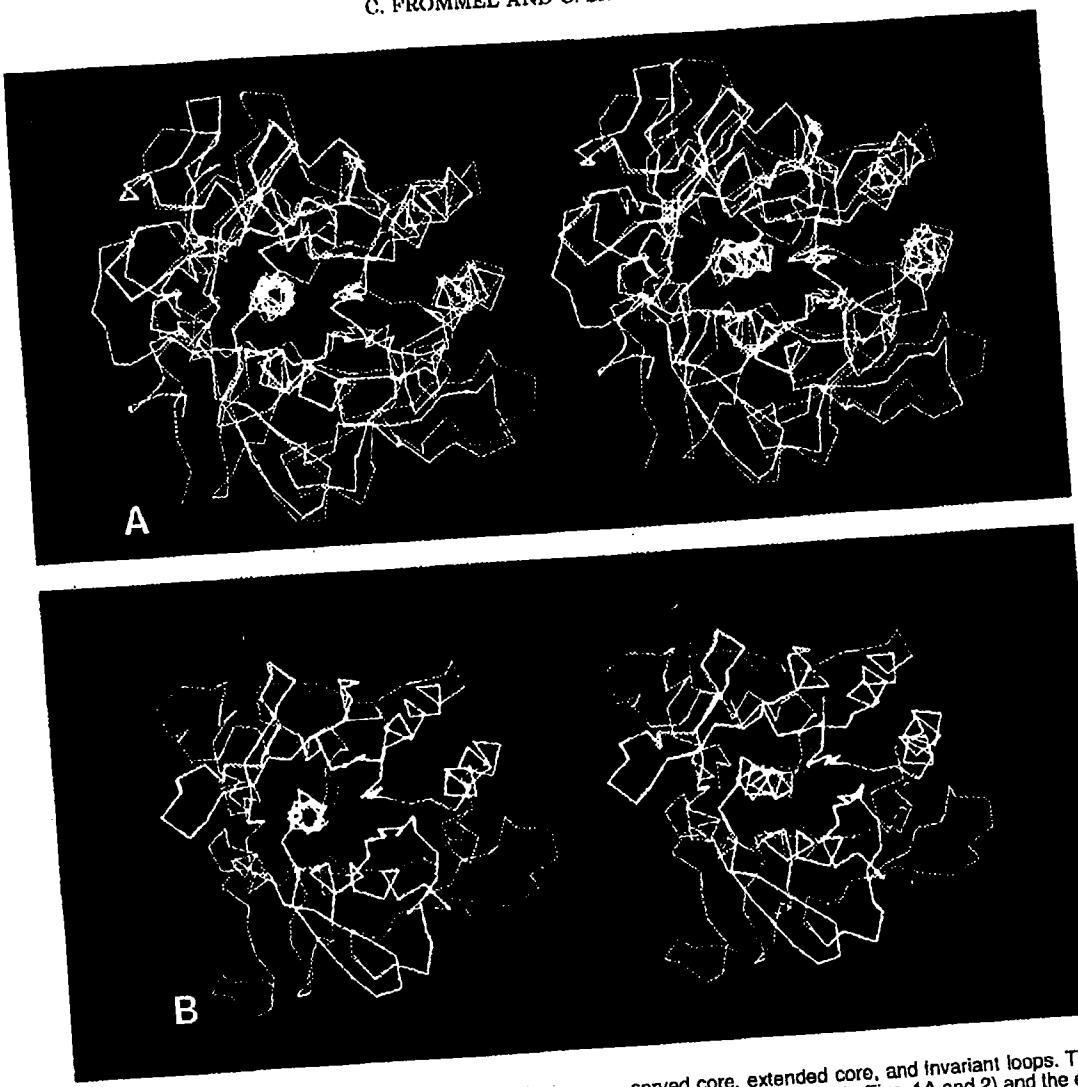


Fig. 1. **A:** Shows the basis of model building by homology. The three-dimensional topology is conserved between the crystal structures of SBT (orange/red) and PRK (green), in spite of a 63% difference in amino acid sequence. Shows also the way in which loop regions (orange and green traces diverge) vary outside of the structural core (orange and green traces nearly coincide). The view is such that the central helices, which carry the active site Ser, are optimally superposed.[58] The rms deviation SBT/PRK of α carbon positions in the conserved core is 0.7 Å. Only the residue centers, i.e., C(α) atoms, are traced. **B:** Shows the three structural layers used in the construction of the initial TRM model: conserved core, extended core, and invariant loops. The conserved core (yellow, defined as in Figs. 1A and 2) and the extended core (violet, some sequence agreement between TRM and SBT) are taken from the known SBT structure. In areas of strong sequence difference, like insertions and deletions, the variant loops of SBT (blue) are discarded in favor of TRM-specific model loops (red). Top of picture: four residue deletion in going from SBT (long blue loop) to TRM (short red loop). Bottom of picture: dangling seven residue insertion in TRM (red) at N terminus. Model regularization by FRODO.[59] Subtilisin coordinates from 1SBT (Protein Data Bank).

ters of polar uncharged groups are well optimized, the electrostatics of interaction of charged side chains (salt bridges etc.) is not well handled by present force field methods (see below). One important reason for this shortcoming is the omission of solvent effects. Explicit simulation of water molecules or inclusion of an effective protein–water potential would correct some of these deficiencies.

Accuracy of Prediction of Loop Conformations

In detail, how good are the predictions of the seven model-built loop regions (Tables II and III, Fig. 3B) in thermitase? Comparing the energy-minimized first model (TRM1) with the preliminary X-ray structure we find three loops with a C(α) rms deviation of less than 1.5 Å (Table II). The deviation is calculated in the context of a best superposition of adjacent core segments. Two loops with deletions of two and four residues are also quite correct. The remaining two loops are incorrect [Table III, deviations in C(α) position relative to the starting conformation is more than 3 Å for three residues]. The bad predictions occur in loops where the sequence alignment on which the model was based has an insertion/deletion shifted in position relative to the inser-

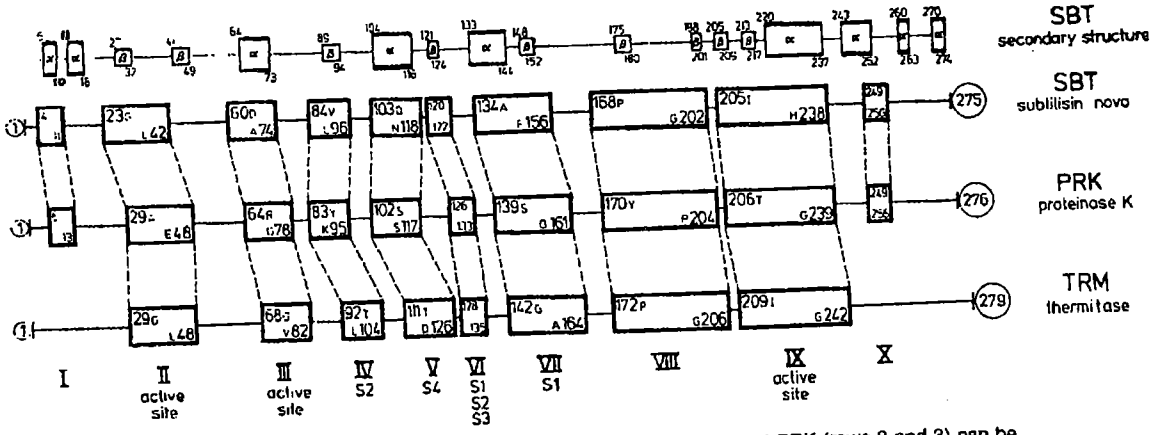


Fig. 2. Segments conserved in the crystal structures of SBT and PRK (rows 2 and 3) can be mapped onto TRM (row 4) by sequence comparison and become the starting core for model building. Structurally conserved segments tend to encompass helices and strands of secondary structure (row 1).

tion/deletion of the structural alignment. Note that none of the methods used here (energy minimization, molecular dynamics) was capable of shifting from the starting conformation sufficiently to correct for the misalignment of sequences or of giving evidence that the starting conformation was wrong in these parts.

Molecular Causes of Thermostability as Deduced from Comparison of Three-Dimensional Structures

A detailed understanding of the molecular reasons for the thermostability of thermitase, even in a simple kinetic model involving one rate-limiting step to irreversible denaturation, requires knowledge of details of interactions in both the native state and the activated state. In other words, understanding the difference in the rates of denaturation between SBT and TRM requires estimating a difference of differences. In what follows we attempt to estimate the difference in the *native* state free energy levels of TRM and SBT relative to the reference state of a completely unfolded protein by analyzing details of molecular interactions given the *native* state coordinates. Differences in the *activated* state free energy levels of TRM and SBT are inaccessible to us, as we have no molecular model for the activated states, unless we make the assumption (as we do here) that the activated states are equally distant in free energy from the completely unfolded reference state for the two proteins. However, any inferences drawn in this way about the causes of thermal stability become invalid if there are significant structural differences in the activated states of TRM and SBT.

Our analysis of hydrogen bond energies, other nonbonded energies, and solvation energies of polar groups at the protein surface sheds little light on the reasons for thermostability, in part because a highly resolved three-dimensional structure is not yet available for TRM. Changes in hydrogen bonds as subtle as those in a mutated subtilisin[6] can of course not be identified in TRM at this stage of structure determination.

Salt bridges

Qualitatively, there is little doubt that salt bridges are important for the stability and folding of some proteins, with detailed examples provided by X-ray crystallography.[32-35] Quantitatively, however, the estimation of interaction energy between charged residues remains difficult in spite of some progress in the theory of protein electrostatics, especially in calculations including the effects of polarizability and solvent screening.[36-39] Our simple count of salt bridges in the experimental structure ØTRM yields an increase of two in the number of salt bridges relative to SCBØ (Table IV, Fig. 5A–D). All salt bridges are conserved between TRM and SCB are accessible to solvent in both structures. Consequently their strength can be influenced by a change in solvent medium (see below).

Aromatic rings

A recent hypothesis, backed by quantum mechanical calculations,[19] attributes as much as 1 kcal/mol to one close (perpendicular) ring–ring interaction of prolines. There are more (n = 24) ring side chains in TRM (in the experimental structure three aromatic rings are not included in analysis) than in SBT (n = 16) or in SCB (n = 18). However, the number of ring–ring contacts in TRM (n = 7) and SCB (n = 6) is comparable while the stability of TRM and SCB is not; the converse is true for SCB (n = 6) and SBT (n = 4)—different number of ring–ring contacts, similar stability (Table IV, Fig. 5E–H). These facts

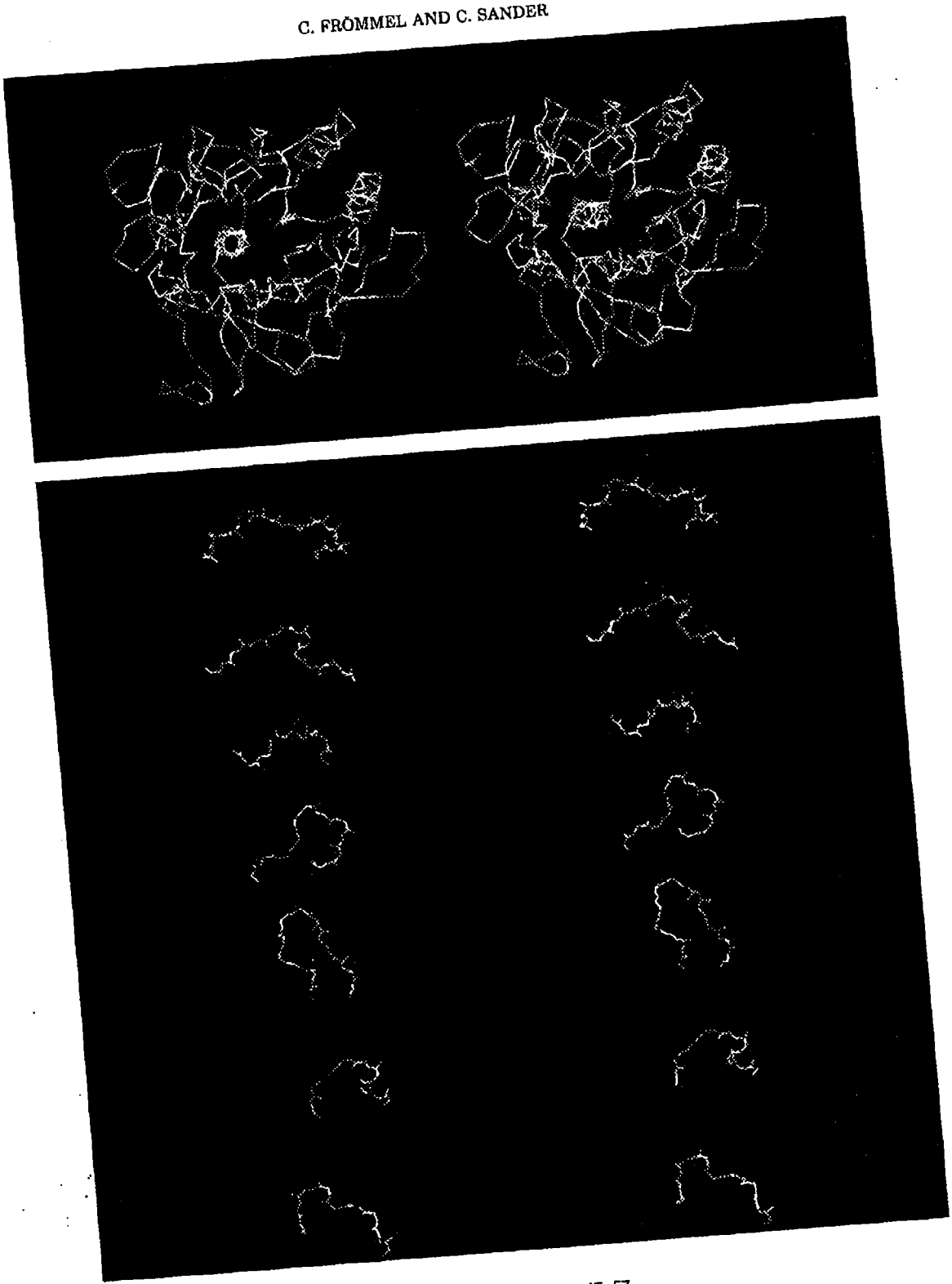


Fig. 3. **A:** C(α) trace of the superposition of the experimental[9] and predicted structure of thermitase shows good agreement except in some of the loop regions. Green: model TRM1. Red: preliminary X-ray structure ([9] and A.V. Teplyakov, personal communication). **B:** Backbone details of the superposition of the experimental[9] and predicted structure of thermitase in seven loop regions, showing the variation in the quality of prediction. From top to bottom:
L1: 24–34
L2: 47–57
L3: 60–68
L4: 79–89
L5: 160–172
L6: 237–245
L7: 255–262
Green: model TRM1. Red: preliminary X-ray structure.[9] See Tables II and III for numerical details. Note that loops L1, L5, and L7 are shorter in this figure than in Table II, for reasons of clarity.

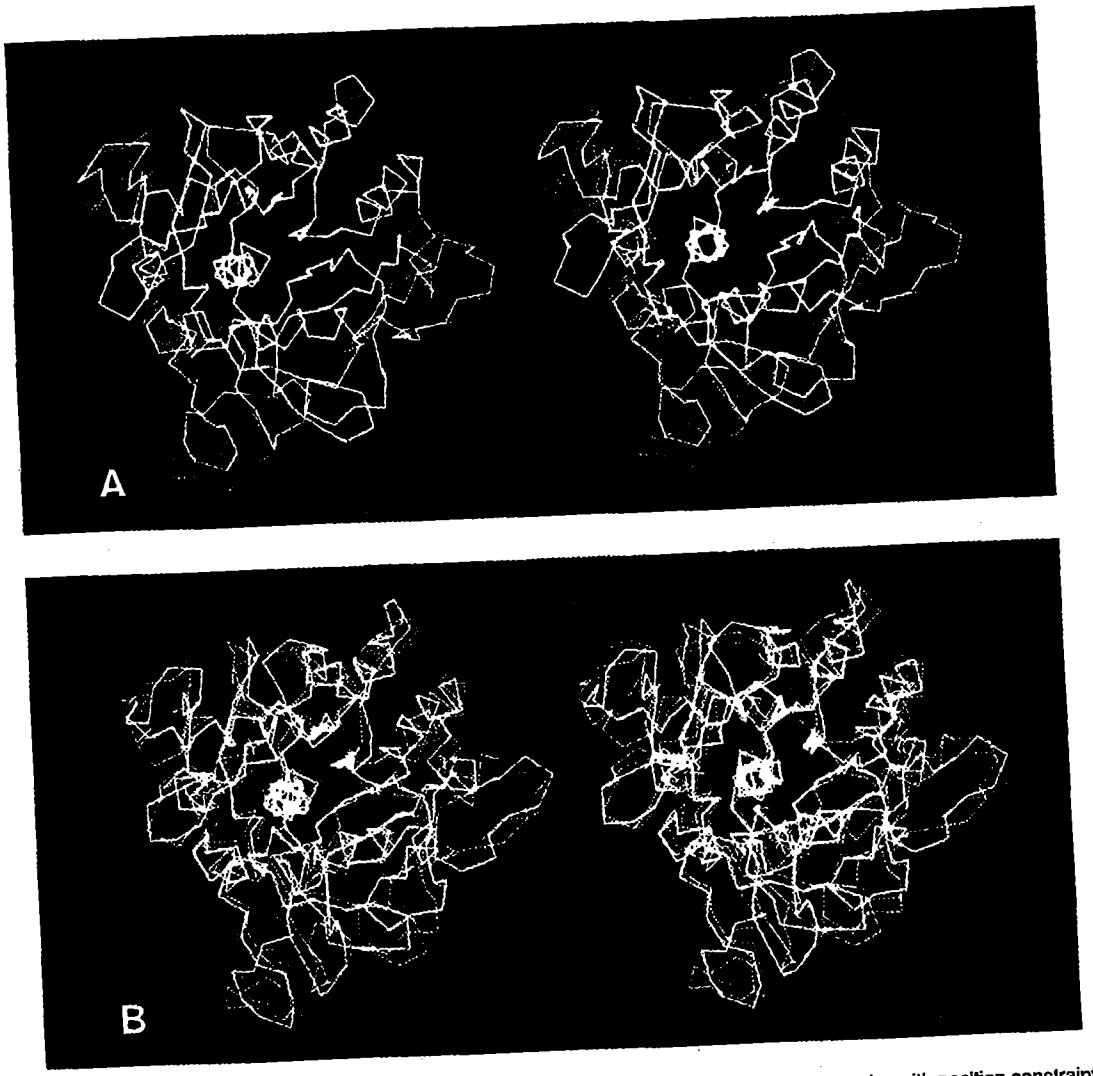


Fig. 4. Refinement of the starting model of TRM by energy minimization (red) and variation by molecular dynamics (green, blue), using GROMOS.[17] The 1050 cycle energy minimized structure (red) differs from the starting model (Fig. 1B) by 0.3 Å rms of C(α) positions. Molecular dynamics at 300° K for several psec samples the allowed conformational space near the starting model. The white parts of the chain (superposition of red, green, and blue!) have not moved during the simulation. **A:** Structures generated by molecular dynamics with position constraints on the Cα atoms of the extended core as defined in Fig. 1B, i.e., only loops outside of the core are allowed to move. Largest C(α) displacement 5 Å. Blue, snapshot at 2.5 psec, minimized. Green, snapshot at 7.5 psec, minimized. **B:** Structures generated by molecular dynamics without any position constraints: all atoms are allowed to move. Largest C(α) displacement 6 Å. Blue, snapshot at 2.5 psec, minimized. Green, snapshot at 7.5 psec, minimized.

appear incompatible with a special role of ring–ring contacts in the stabilization of TRM.

Hydrophobic interaction

We use the relation between apolar surface area and transfer energy[20] to estimate the hydrophobic contribution to the folded/unfolded free energy difference. Between the three experimental structures (TRM, SBT, and SCB) there is no significant difference in calculated hydrophobic stabilization; about -140 ± 7 kcal/mol for the entire protein molecule, -99 ± 3 kcal/mol for the conserved core, and -77 ± 2 kcal/mol for the buried parts of the core. Experiments (see below) are consistent with this estimate.

Ca binding

There is crystallographic evidence for two binding sites in the SCB and SBT structures.[52] The *weaker* of the two sites (dissociation constant 10^{-4} M, Ca easily removable by EDTA) involves residues Ala-168, Lys-169, Tyr-170, Val-173, and Glu-194 in SCB and Gly-169, Lys-170, Tyr-171, Val-174, and Asp-195 in SBT. As we do not yet have crystallographic data on Ca binding in TRM, we identify this TRM site by homology as involving Ala-173, Tyr-174, Tyr-175, Ala-178, and Asp-200. The *stronger* of the two sites (dissociation constant 10^{-8} M[52] involves Gln-2, Asp-41, Leu-75, and Asn-77 in SBT and the correspond-

TABLE II. The Seven Critical Loops in the Tertiary Structure Prediction of Thermitase: Deviation of Five Model Structures from the Experimental Structure*

Loop region	Models					Remarks
	TRMØ	TRM1	TRM2	TRM3	TRM4	
21–34	1.64	1.46	3.60	1.41	2.82	Deletion of one residue
47–57	2.75	2.69	3.05	2.23	3.00	Insertion of two residues, Sequence alignment off
60–68	1.97	2.00	3.03	2.15	2.96	Deletion of one residue, sequence alignment off
79–89	1.44	1.34	1.68	1.93	1.75	Side chain orientation incorrect
160–177	1.97	1.95	2.50	2.88	2.62	Deletion of four residues
237–245	1.75	1.62	0.80	1.41	1.38	Deletion of two residues
254–269	1.36	1.39	2.31	1.82	2.28	Insertion of one residue

*rms, deviation of C(α) positions.
TRMØ, model after geometric regularization.
TRM1, energy minimized TRMØ (1050 cycles).
TRM2, snapshot after 7.5 psec molecular dynamics and energy minimization.
TRM3, snapshot after 7.5 psec molecular dynamics with constraints on C(α) positions of the core structure (defined in Table I) and energy minimization.
TRM4, structure averaged over molecular dynamics trajectory between 5.0 and 7.5 psec.
The N-terminal 7 amino acids are omitted here as they are not visible in the electron density.

ing conserved (Table IA) residues in SCB. In TRM this site appears to be well conserved, except for the replacement of Gln-2 by Ser. As we have no theoretically reliable way of calculating the influence of Ca binding on stability against denaturation, the tentative identification of these two Ca binding sites in the TRM structure in itself sheds no light on the causes of thermostability; however, in conjunction with experimental binding studies (see below), identification of these sites leads to an important hypothesis about a third site.

Molecular Causes of Thermostability as Deduced from Stability and Binding Experiments

The stability of irreversibly denaturating proteases can be increased either by strengthening stabilizing interactions in the ground state (native, folded state) or by weakening stabilizing interactions in the activated state of denaturation/autolysis. The activated state is not yet accessible to experiment nor to model building, but we can obtain additional independent evidence about the influence of particular interactions on the difference between the ground state and the activated state by measuring the rates of inactivation under conditions where one of these interactions is thought to be strengthened.

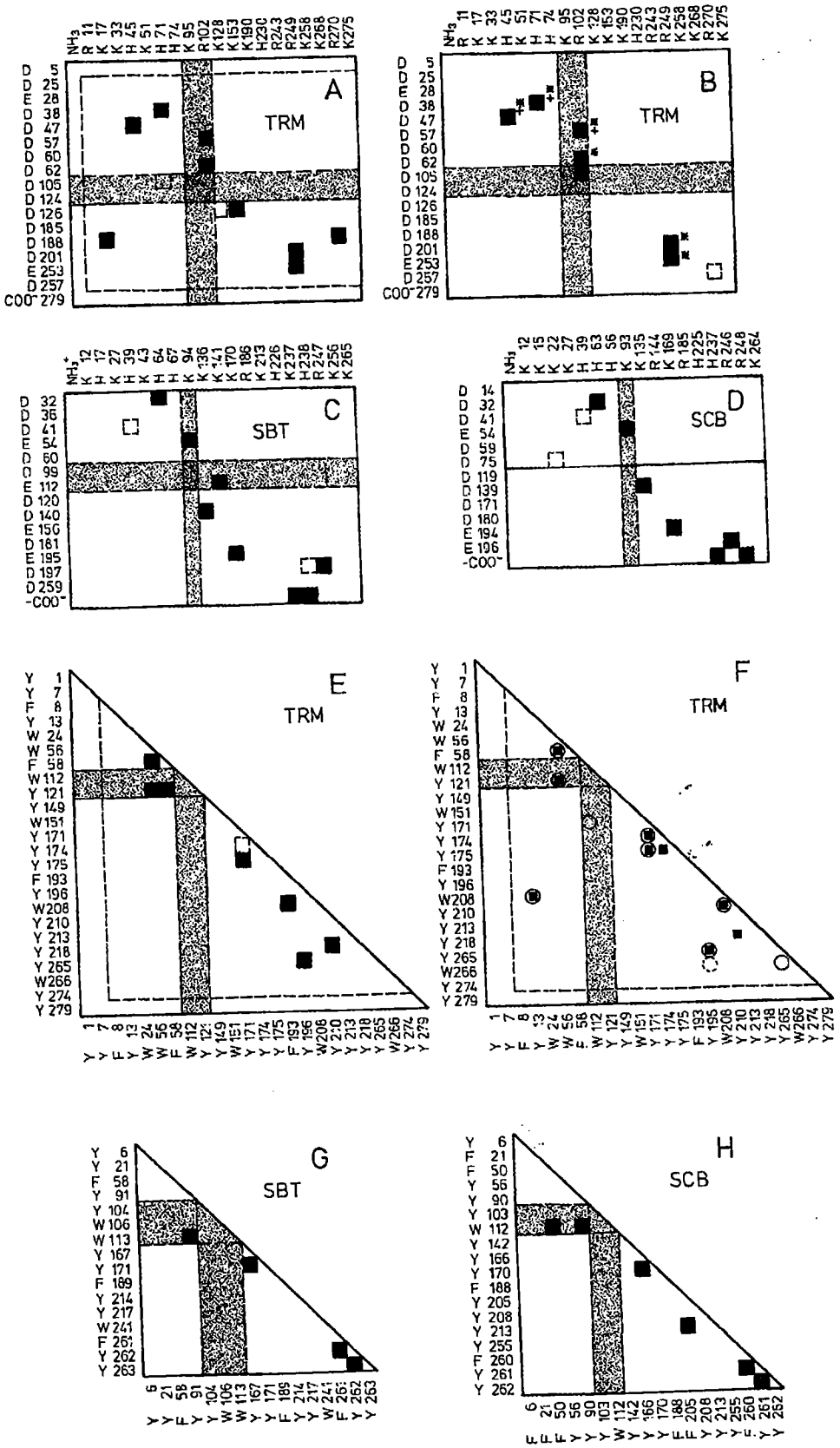
Salt bridges

The attempt to confirm by experiment a structural role for the few additional salt bridges in TRM by varying the ionic strength failed to show even a small difference between TRM and SCB. A variety of salts and 0.1 M Na_2SO_4 all have a weakly stabilizing influence on both TRM and SCB, with no dependence on the type of salt used (Fig. 6). If the few additional salt bridges in TRM were important for stability there would be an influence of ionic strength on thermostability as none of the salt

Fig. 5. Are a few localized interactions responsible for thermostability? Comparison of residue–residue interactions in the known structures of SCB and SBT with those in TRM (preliminary experimental data and model structure). The shaded strips mark the hinge residues between the two structural domains.[60] TRM residues outside of the broken line square are ignored (missing in X-ray structure). Residues are given in one letter code with their Protein Data Bank sequence number.

Rectangular graphs—interaction between ionizable groups: An interaction is marked if atoms which can carry charge are within a cutoff distance of 4.0 Å (black squares) or between 4.0 and 5.0 Å (dotted squares). **A:** TRM experimental structure. **B:** TRM model structures: squares, TRM1 (energy minimized, 1050 cycles); stars, interactions correctly predicted in TRM1; plusses, TRM3 (after 7.5 psec molecular dynamics and energy minimization). **C:** Subtilisin BPN'. **D:** Subtilisin Carlsberg. The nearly buried salt bridges (less than 20% of either partner accessible) in SBT are D32/H64 (active site), H39/D41, K94/D54, and K170/E195. Of these, three are conserved in TRM as D38/H71, H45/D47, and R102/D57, one is a hydrophobic contact in TRM: Y174/V199. Note that 1–7 (YTPNDPY) and 279 (Y) in TRM are not included and that in SBT and SCB the carboxylic group of the C-terminus participates in two ionic interactions; consequently the complete structure of TRM may have more salt bridges than indicated here. A remarkable difference between TRM and the subtilisins is the interdomain salt bridge (K17, D188), accessible to water (homologous residues in SBT are Q10/N188 and in SCB L10/S187—no salt bridge). There are no contacts between equally charged groups closer than 6 Å.

Triangular graphs—ring–ring interactions: Interactions between Phe, Tyr, and Trp. Distance cutoff at 7.0 (Phe-Phe), 8.5 (Trp-Trp), and 8.0 (Trp-Phe) Å. Distances are between the centers of mass of rings. **E:** TRM experimental structure. **F:** TRM model structures: squares, TRM1 (energy minimized, 1050 cycles); circles, TRM3 (after 7.5 psec molecular dynamics and energy minimization). **G:** subtilisin BPN'. **H:** Subtilisin Carlsberg



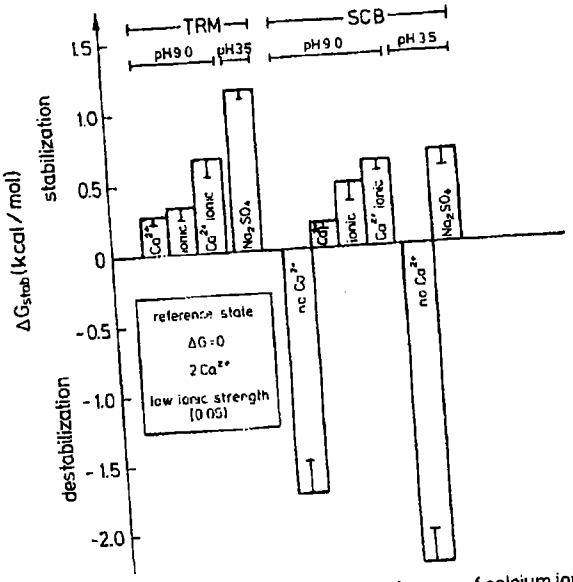


Fig. 6. Experimental evidence for the influence of calcium ions and different salts on the stability of thermitase and subtilisin Carlsberg at 60°C. Ca: stabilization due to adding free calcium ions at a concentration of 1.0 mM; no Ca: destabilization due to removal of Ca from the enzyme by EDTA; does not work for the Ca ions very tightly bound to TRM; ionic: stabilization due to adding any one of the salts NaCl, KCl, $NaCH_3COO$, KCH_3COO, $(NH_4)_2SO_4$, or Na_2SO_4 at an ionic strength of $I = 0.25$ to a reference buffer solution with $I = 0.05$; Na_2SO_4: stabilization due to adding 1.0 M sodium sulfate; pH 9.0: borate buffer, $I = 0.05$, pH = 9.0; pH 3.5: acetate/borate buffer, $I = 0.05$, pH = 3.5. Enzyme concentration was $2.0-5.0 \times 10^{-5}$ M. The stabilization energy was estimated from inactivation rate constants k by $G(\text{stab}) = RT \ln[k(0)/k]$, where (0) refers to the reference state. All values are the mean of at least five estimates with the standard deviation marked inside each bar. Reference state: TRM, two bound calcium ions per enzyme molecule; SCB, one bound calcium ion, at low ionic strength ($I = 0.05$). In the reference state, TRM is more stable than SCB by 0.75 kcal/mol at pH 9.0 and by 1.39 kcal/mol at pH 3.5 (data not shown). Complete removal of Ca from SCB is achieved in 30 minutes by EDTA (Frömmel and Briedigkeit, unpublished, and [52]). Due to the low ionic product of calcium ions and sulfate (10^{-5}), the stabilizing effect of sulfates could only be estimated without the third weakly bound calcium ion.[61]

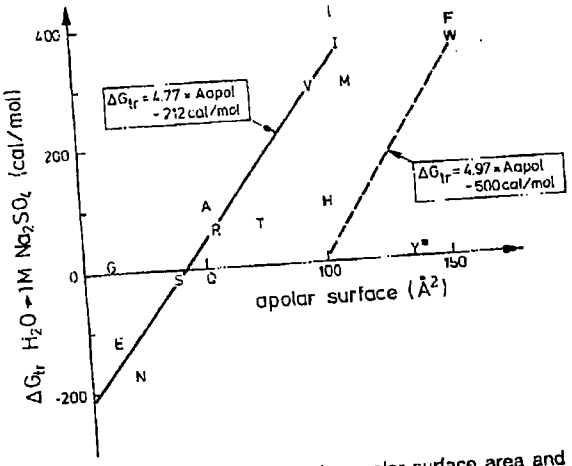


Fig. 7. Correlation between the apolar surface area and the free energy of transfer for amino acid side chains is the basis for estimating (Fig. 6) changes in hydrophobic interactions during the inactivation process. The apolar surface area is defined as in Ref. (20). ΔG_{tr} is from transfer experiments between water and 1 M sodium sulfate solution.[41] The correlation coefficients are $r = 0.94$ for the 13 amino acids without Tyr, Phe, Trp, and His and $r = 0.81$ for these four amino acids. Tyr (*) has unusually low solubility and a large experimental error. Amino acids are given in one letter code.

TABLE III. Mismatches in Sequence Alignment Leading to Incorrect Structure Prediction*

```
alignment based on sequence

SBT     41 - 49        D L - - K V A G G A S
                       X X     X X   X X
TRM     47 - 57        D L A G K V V G G W D

alignment based on experimental 3-D structure

SBT     41 - 49        D L K V A   G - G   A S
                       X X       / \ / \
0TRM    47 - 57        D L A G K V V G G W D

alignment based on sequence

SBT     53 - 61        S E T N P F Q D D
                       : : : : X   X : :
TRM     61 - 68        N D S T P - Q N N

alignment based on experimental 3-D structure

SBT     53 - 61        S E T N P F Q D D
                       : : \ /     X : :
0TRM    61 - 68        N D   S   T P Q N N
```

*X, identity; :, similarity; /\, ambiguity. 0TRM, preliminary crystal structure of thermitase.[9]

bridges is inaccessible to solvent in the native structure. As such an influence is not observed we conclude that salt bridges play no role in stabilizing either the ground state or destabilizing the activated state of thermitase.

Hydrophobic stabilization

Measurement of denaturation rates with and without Na_2SO_4 can provide evidence for the role of hydrophobic stabilization if one uses the standard hypothesis that Na_2SO_4 stabilizes protein structure solely by decreasing the water solubility of the more hydrophobic side chains.[40,41] The result is that TRM is stabilized relative to SCB by hydrophobic interactions by about 0.5 kcal/mol (see difference between the two bars labeled Na_2SO_4 in Fig. 6). The implication is that TRM exposes more hydrophobic groups to solven than does SCB in going from the ground to the activated state.

In order to estimate how many hydrophobic groups may be involved, we exploit a correlation be-

TABLE IV. Comparison of Predicted and Observed Structures: Surfaces, Main Chain Hydrogen Bonds, Salt Bridges, and Ring–Ring Interactions*

	ØTRM exp†	TRMØ mod	TRM1 mod	TRM2 mod	TRM3 mod	SCBØ exp	SBTØ exp	SBT1 min	SBT2 md
Nonbonded energy [10^3 kcal/mol]	−3.2	> +10000	−3.0	−3.4	−3.3	−2.5	−2.0	−3.4	−3.3
Surface area (accessible, % of maximal)									
Total	28.1	29.4	29.1	27.2	28.9	25.4	25.5	25.4	25.0
Polar	29.1	29.7	29.6	25.6	27.3	26.0	25.6	25.6	23.6
Apolar	26.7	29.1	28.1	29.2	31.1	24.6	25.1	25.2	26.8
Main chain H-bonds per 100 residues (according to DSSP[57])									
Helix	50‡	65.9	67.7	76.3	69.5	68.2	69.8	70.2	78.5
All	109	128.3	139.1	149.5	139.1	130.0	133.1	133.1	148.4
Salt bridges (distance ≤ 4 Å)									
From coordinates	9	7	7	0	3	7	8	8	6
Correct	—	6	6	0	3	—	—	8	4
Not found	—	3	3	9	6	—	—	0	4
Ring–ring interactions§									
From coordinates	7	9	9	9	9	6	4	4	4
Correct	—	4	4	5	6	—	—	4	3
Not found	—	3	3	2	1	—	—	0	1
	exp	mod	mod	mod	mod	exp	exp	min	md

*exp, experimental structure from X-ray crystallography.
mod, model structure (this paper).
ØTRM, crystallographic structure of thermitase[9]; energy after energy minimization for 200 cycles.
TRMØ, geometrically optimized model of thermitase, starting structure.
TRM1, energy minimized model (1050 cycles), starting from TRM0.
TRM2, model of thermitase after 7.5 psec molecular dynamics and energy minimization, starting from TRM1.
TRM3, model of thermitase after 7.5 psec molecular dynamics with constraints on 166 C(α) positions in the conserved core (defined as in Table I) and energy minimization, starting from TRM1.
SBTØ, subtilisin BPN′ (Mike James and Catherine A. McPhalen, prepublication gift).
SCBØ, subtilisin Carlsberg (Mike James and Catherine A. McPhalen, prepublication gift); subsequent energy minimization (GROMOS) makes very small changes in the structure and decreases the energy to -3.3×10^3 kcal/mol.
SBT1, energy-minimized (min) structure (250 cycles) starting from SBTØ.
SBT2, structure of subtilisin after 7.5 psec molecular dynamics (md) starting from SBT1, energy minimized.
†Excluding residues 1–7 and 279 (not in the preliminary X-ray structure).
‡Due to the incomplete refinement of the thermitase X-ray coordinates, secondary structure assignments by DSSP[57] are rather conservative, i.e., helices and strands are 1–3 residues shorter than would be implied by the homology to the subtilisin structures.
§Ring–Ring interaction is counted if distance of side chain center of mass Phe-Phe, Tyr-Tyr <7.0 Å, Phe-Trp, Tyr-Trp < 8.0 Å, and Trp-Trp < 8.5 Å.
*The different types of accessible surface area are defined as in Ref. (20).

tween a quantity called apolar surface introduced by Frömmel[20] and Bull and Swenson's data[41] for the transfer of amino acids from water to Na_2SO_4 (Fig. 7). For a folded protein this correlation translates into a rule that uncovering 1 Å^2 of apolar surface area is more difficult by 0.005 kcal/mol in 1 M Na_2SO_4 than it is in pure water. Using this rule, we deduce that the measured additional kinetic stabilization of TRM by the hydrophobic effect corresponds to a difference in apolar surface area of about 100 Å^2. This small amount of surface area corresponds to a few methyl side chain groups (note that Ala as a polypeptide residue has a maximal exposed surface of 115 Å^2).

In summary, there appears to be some hydrophobic stabilization of TRM over SCB. However, the effect is too small to explain the 20-fold higher thermostability of thermitase with 1 mM EDTA present (see difference in bars labeled "no Ca^{2+}" in Fig. 6).

Ca binding

Several known[10] and new experimental facts point to a stabilizing role of calcium. (1) In addition to the two binding sites identified in the TRM structure by homology to know sites in SCB and SBT, there is a third site (measurement of absolute Ca content by atomic absorption spectroscopy[10]). (2) Two of the three Ca ions in TRM are bound more tightly than in subtilisin; the third Ca site in TRM corresponds to the weakly bound Ca in SCB.[10] The stabilizing effect of the weakly bound Ca is comparable in SCB and TRM (bars labeled Ca^{2+} in Fig. 6). (3) The complete removal of Ca ions from subtilisin leads to a significant decrease in stability (Fig. 6). (4) It has so far been impossible to completely remove bound Ca ions without at the same time destroying the thermitase structure. The presence of 0.1 M EDTA reduces the number of bound Ca ions

TABLE V. The Role of Particular Interactions in Thermitase Compared to Subtilisin Carlsberg: Predictions and Experiments

	Theoretical analysis of interactions			Experimental observations
	TRM model minimized*	TRM model after dynamics*	TRM X-ray structure*	Stability measurements on TRM and SCB
Hydrophobic stabilization	Similar in the core, lower in loops	Similar in the core, much lower in the loops	No significant difference	Small kinetic stabilization of TRM relative to SCB, due to an increased difference in hydrophobic interaction between ground state and activated state
Number of salt bridges	Slightly higher, extra salt bridges are water accessible	Lower	Extra bridges (water accessible); one extra bridge between domains†	The stability of both TRM and SCB is increased if electrostatic interactions are weakened by addition of salt (Fig. 5.)
Number of ring–ring contacts	Higher	Higher	Slightly higher†	Stability of SCB and SBT similar[26] in spite of difference in the number of ring–ring interactions (6 for SCB, 4 for SBT)
Calcium binding	Two sites conserved	Two sites conserved	[No X-ray data on Ca binding sites in TRM]	Binding site common to all subtilisins with dissociation constant of 0.1 mM; one additional tightly bound Ca ion in SCB, two additional in TRM, for a total of 3 bound Ca ions in TRM

*Changes in TRM relative to SCB, e.g., "lower" means TRM has lower number than SCB.
†Not included in these counts are the first seven (YTPNDPY) and the last residue (Y) of TRM, which are not visible in the electron density map.

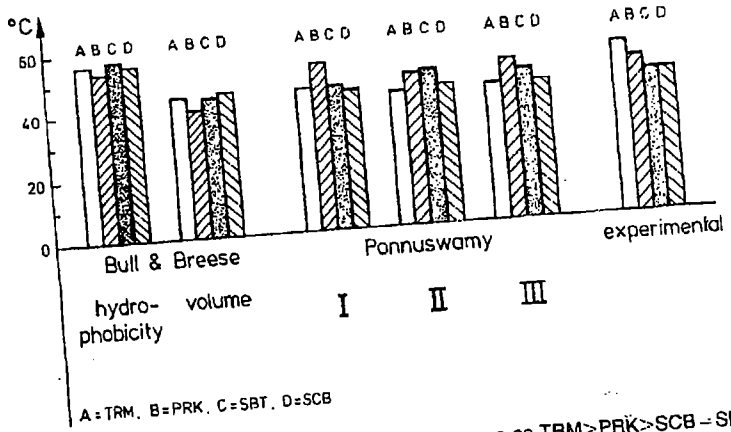


Fig. 8. Temperature of unfolding of subtilisin proteases. **A:** TRM, **B:** PRK, **C:** SBT; **D:** SCB. Agreement between (right) predictions by various methods (left and middle) and experiment is poor. The experimentally obtained stability of the proteases decreases as TRM>PRK>SCB ≈ SBT.[10,26] Predictions were made on the basis of amino acid composition weighted by hydrophobicity or volume of the residues[43] or by use of three different groupings of amino acids.[47]

per TRM protein molecule from three to one within 2 hours, but further incubation for another 15 days leads to no further appreciable dissociation. In the absence of other convincing factors (Table 5), we take the large destabilizing effect of the removal of Ca ions on subtilisin and the unusually strong binding of at least one Ca ion in thermitase as evidence for a major role of Ca binding in the unusual thermostability of TRM.

The additional strong Ca binding site remains to be identified in the structure of TRM. We predict this site to involve oxygens from side chains in a loop near Asp-57, namely Asp-60, Asp-62, Gln-66, Asn-67, and/or Asn-69. This loop faces Asp-105 which could participate in the hypothetical calcium binding site. The accuracy of this prediction can be assessed when bound Ca becomes visible in the refined crystal structure.

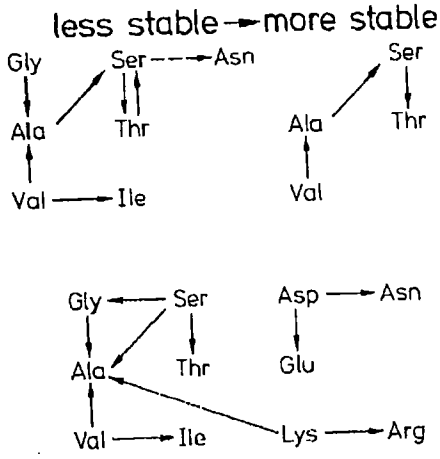


Fig. 9. Amino acid exchanges observed in the transition to higher thermostability. Subtilisin proteases top left. Conserved core of subtilisin proteases top right. Dehydrogenases and ferredoxins[48] bottom. Only the most significant[48] exchanges are shown. Based on comparison of TRM with SCB, SBT, and PRK. The control, comparing SCB, SBT, and PRK with one another, shows no significant exchanges.

Theories of the Origin of Thermostability

Is there a general theory of thermostability applicable to thermitase? Can our arguments for stability in the subtilisin family be generalized to other protein classes? A number of labile–stable pairs of homologous enzymes from different organisms are known. Typically, the two proteins in the pair are more than 50% homologous in sequence and clearly homologous in the topology of tertiary structure, but differ in thermostability by tens of °C. Based on empirical analysis of these labile–stable pairs, several hypotheses have been proposed in an attempt to qualitatively explain the difference in stability or to quantitatively predict the temperature of unfolding.

Amino acid content

One type of theory explains thermostability by stating that certain of the 20 standard residues are used preferentially in more stable proteins[42–47] without, however, providing a physical rationale for the effect. For example, Bull and Breese[43] weight the amino acid content of proteins by hydrophobicity and volume. Ponnuswamy et al.[47] correlate directly the amino acid content of proteins with their thermostability, and Ikai[44] with an aliphatic index. Upon application of any of these rules to the subtilisin proteases, we fail to see agreement between the calculated and predicted temperatures of unfolding (Fig. 8). Neither the absolute value nor the relative stability among the members of this family is correctly predicted. For example, the value of the aliphatic index[44] is expected to increase with increasing thermostability, but in fact does not (the index is 71.8, 62.1, 84.8, 84.8 for TRM, PRK, SBT, SCB, respectively).

Changes in amino acid content

Another class of analyses reports the preferred amino acid residue content of thermostable proteins in terms of evolutionary amino acid replacements, typically as a table of replacement pairs, like Ala to Val, derived from known labile–stable pairs of proteins. Some of the statistically preferred stabilizing replacements in the subtilisin family (Fig. 9) agree with those derived from the family of dehydrogenases and ferredoxins[48]: Val-to-Ala, Gly-to-Ala, Val-to-Ile; others disagree, e.g., Ala-Ser. In the structurally conserved core of subtilisin, the trends are Val-to-Ala, Ala-to-Ser, and Ser-to-Thr.

Single point mutations

A number of different hypotheses follow from point mutation experiments aimed at improving thermostability.[6,49,50] For example, the single exchange of Asn-218 to Ser in subtilisin appears to lead to a decrease in the rate of inactivation by a factor of 4, either due to a decrease in activity, resulting in reduced autolysis, or due to a decrease in the rate of unfolding.[6] The high-resolution difference electron density map suggests a subtle improvement in the hydrogen bonding scheme near a beta bulge. Other examples of stabilizing mutations in tryptophan synthase[50] and human hemoglobin[51] are also difficult to interpret. However, these experiments do show that very localized interactions can lead to a significant increase in stability against unfolding.

In terms of these two views—average drift in amino acid content versus changes in localized interactions—the results of our analysis (Table 5) tend to agree with the second view. In the subtilisin family, the stability of thermitase appears to be due to localized changes involving one or two Ca binding sites.

Additional supporting evidence for our hypothesis may come from locating Ca ions in the refined crystal structure, when it becomes available. A direct experimental test of our Ca binding hypothesis would become possible if we were able to remove the firmly bound Ca ions from thermitase while maintaining function. Perhaps the most conclusive experiment would involve elimination of the strongest Ca binding site from thermitase by point mutations.

ACKNOWLEDGMENTS

We would like to thank Mike James and Catherine McPhalen, Edmonton, Canada, for prepublication subtilisin coordinates; Wolfram Sänger, Berlin (West), for C(α) coordinates of proteinase K; Wilfred van Gunsteren and Herman Berendsen, Groningen, for use of the molecular dynamics software GROMOS; Alwyn Jones, Uppsala, for use of the graphics software FRODO; at EMBL, Heidelberg, Heinz Bosshard and Anna Tramontano for help with

graphics and Johan Postma for help with GROMOS; at MPIMF, Heidelberg, Gabi Eulefeld for figures and photography; the Max-Planck-Gesellschaft for a fellowship; and the Abteilung Biophysik at MPIMF and Ken Holmes for hospitality.

NOTE ADDED IN PROOF

The recently refined crystal structure of thermitase (A.V. Teplyakov, personal communication) has a bound Ca near Asp-57, as predicted here.

REFERENCES

1. Hartley, B.S. Commercial prospects for enzyme engineering. Phil. Trans. R. Soc. London A 317:321–331, 1986.
2. Küttner, G.A.S., Bürger, E., Pfüller, B., Frömmel, C. Investigations on the long-term stability of enzymes in solution. Biomed. Biochim. Acta 46:39–52, 1987.
3. Estell, D.A., Graycar, T.P., Wells, J.A. Engineering an enzyme by site-directed mutagenesis to be resistant to chemical oxidation. J. Biol. Chem. 260:6518–6521, 1985.
4. Wells, J.A., Powers, D.B. *In vivo* formation and stability of engineered disulfide bonds in subtilisin. J. Biol. Chem. 261:6564–6570, 1986.
5. Pantoliano, M.W., Ladner, R.C., Bryan, P.N., Rollence, M.L., Wood, J.F., Poulos, T.L. Protein engineering of subtilisin BPN': Enhanced stabilization through the introduction of two cysteines to form disulfide bond. Biochemistry 26:2077–2082, 1987.
6. Bryan, P.N., Rollence, M.L., Pantoliano, M.W., Wood, J., Finzel, B.C., Gilliland, G.L., Howard, A.J. and Poulos, T.L. Proteases of enhanced stability: Characterization of a thermostable variant of subtilisin. Proteins 1:326–334, 1986.
7. Frömmel, C., Hausdorf, G., Höhne, W.E., Behnke, U., Rutloff, H. Thermitase—a thermostable serine protease from *Thermoactinomyces vulgaris:* One step purification and characterization. Acta Biol. Med. Germ. 37:1193–1204, 1978.
8. Meloun, B., Baudys, M., Kostka, V., Hausdorf, G., Frömmel, C., Höhne, W.E. Complete primary structure of thermitase from *Thermoactinomyces vulgaris* and its structural features related to the subtilisin-type proteinases. FEBS Lett. 183:195–200, 1985.
9. Teplyakov, A.V., Strokopitov, B.W., Kuranowa, I.P, Popov, A.N., Arutyunjan, E.G., Weinshtein, B.K., Frömmel, C., Höhne, W.E. X-ray analysis of thermitase at 2.5 Å resolution. Kristallografia 31:466–469, 1986. (Russian.)
10. Frömmel, C., Höhne, W.E. Influence of calcium binding on the thermal stability of 'thermitase,' a serine protease from *Thermoactinomyces vulgaris*. Biochim. Biophys. Acta 670:25–31, 1981.
11. Yamada, K., Kometam, K. The change in heat capacity and entropy of troponin C induced by calcium binding. J. Biochem. 92:1505–1517, 1982.
12. Smith, T.F., Waterman, M.S. Identification of common molecular subsequences. J. Mol. Biol. 147:195–197, 1981.
13. McLachlan, A.D. Test for comparing related amino acid sequences. Cytochrome C and cytochrome C551. J. Mol. Biol. 61:409–424, 1971.
14. Bernstein, F.C, Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., Tasumi, M. The Protein Data Bank: A computer-based archival file for macromolecular structures. J. Mol. Biol. 112:535–542, 1977.
15. Paehler, A., Banerjee, A., Dattagupta, J.K., Fujiwara, T., Lidner, K., Pal, G.P., Suck, D., Weber, G., Sänger, W. Three-dimensional structure of fungal proteinase K reveals similarity to bacterial subtilisin. EMBO J. 3:1311–1314, 1984.
16. Greer, J. Model for haptoglobin heavy chain based upon structural homology. Proc. Natl. Acad. Sci. U.S.A. 77:3393–3397, 1980.
17. van Gunsteren, W.F., Berendsen, H.J.C., Colonna, F., Perahia, D., Hollenberg, J.P., Lellouch, D. On searching neighbours in computer simulation of macromolecular systems. J. Comput. Chem. 5:272–279, 1984.
18. Singh, J., Thornton, J.M. The interaction between phenylalanine rings in proteins. FEBS Lett. 191:1–6, 1985.
19. Burley, S.K., Petsko, G.A. Aromatic-aromatic interaction: A mechanism of protein stabilization. Science 299:23–28, 1985.
20. Frömmel, C. The apolar surface area of amino acids and its empirical correlation with hydrophobic free energy. J. Theor. Biol. 111:247–260, 1984.
21. Russell, A.J., Fersht, A.R. Commercial samples of subtilisin BPN'. Nature (London) 321:733–733, 1986.
22. Laemmli, U.K. Cleavage of structural proteins during the assembly of the head of becteriophage T4. Nature (London) 227:680–685, 1970.
23. Mokrasch, L.C., McGilvery, R.W. Purification and properties of fructose-1,6-diphosphatcase. J. Biol. Chem. 221:090–917, 1956.
24. Markland, F.S., Smith, E.L. Subtilisins: Primary structure, chemical and physical properties. In: "The Enzymes" (P.D. Boyer, ed.), Vol. III, 3rd ed. New York: Acadamic Press, 561–608.
25. Imoto, T., Yamada, H., Ueda, T. Unfolding rates of globular proteins determined by kinetics of proteolysis. J. Mol. Biol. 190:647–649, 1986.
26. Voordouw, G., Milo, C., Roche, P.S. Role of bound calcium ions in thermostable, proteolytic enzymes, separation of intrinsic and calcium ion contribution to the kinetic thermal stability. Biochemistry 15:3716–3724, 1976.
27. Sibanda, B.L., Blundell, T., Hobart, P.M., Fogliano, M., Bindra, J.S., Dominy, B.W., Chirgwin, J.M. Computer graphics modelling of human renin. Specifity, catalytic activity and intron exon junction. FEBS Lett. 174:102–110, 1984.
28. Lesk, A.M., Chothia, C. How different amino acid sequences determine similar protein structures: The structure and evolutionary dynamics of the globins. J. Mol. Biol. 136:225–270, 1980.
29. Chothia, C., Lesk, A.M. Evolution of proteins formed by beta-sheets. Plastocyanin and azurin. J. Mol. Biol. 160:309–323, 1982.
30. Fine, R.M., Wang, H., Shenkin, P.S., Yarmush, D.L., and Levinthal, C. Predicting antibody hypervariable loop conformations. II: Minimization and molecular dynamics studies of MCPC603 from many randomly generated loop conformations. Proteins 1:342–362, 1986.
31. Moult, J., James, M.N.G. An algorithm for determining the conformation of polypeptide segments in proteins by systematic search. Proteins 1:146–163, 1986.
32. Walker, J.E., Wonacatt, A.J., Haaris, J.J. Heat stability of a tetrameric enzyme d-glyceraldehyde-3-phosphate dehydrogenase. Eur. J. Biochem. 108:581–586, 1980.
33. Sunderam, T.K., Chell, R.M., Wilkinson, A.E. Monomeric malate synthase from thermophilic bacillus. Molecular and kinetic characteristics. Arch. Biochem. Biophys. 199: 515–525, 1980.
34. Perutz, M.F., Raidt, H. Stereochemical basis of heat stability in bacterial ferredoxin and in haemoglobin A2. Nature (London) 255:256–259, 1975.
35. Warshel, A., Russell, S.T. Calculation of electrostatic interactions in biological systems and in solution. Quart. Rev. Biophys. 17:283–422, 1984.
36. Russell, S.T., Warshel, A. Calculation of electrostatic energies in proteins. The energetics of ionized groups in bovine pancreatic trypsin inhibitor. J. Mol. Biol. 185, 389–404. 1985.
37. Warshel, A., Russell, S.T., Churg, A.K. Macroscopic models for studies of electrostatic interactions in proteins: Limitations and applicability. Proc. Natl. Acad. Sci. U.S.A. 81:4785–4789, 1984.
38. Rashin, A.A., Honig, B. (1984). On the environment of ionizable groups in globular proteins. J. Mol. Biol. 173:515–521, 1984.
39. Honig, B.H., Hubbell, W.L. Stability of 'salt bridges' in membrane proteins. Proc. Natl. Acad. U.S.A. 81:5412–5416, 1984.
40. von Hippel, P.H., Schleich, T. The effects of neutral salts on the structure and conformational stability of macromolecules in solution. In: "Structure and Stability of Biological Macromolecules" (S.N. Timasheff and G.D. Fasman, eds.). New York: Dekker, 1969:417–574.

41. Bull, H.B., Swenson, C. Solubility of amino acids. Biopolymers 17:1091–1100, 1978.
42. Singleton, R., Middaugh, C.R., McElroy, R.P. Comparison of proteins from thermophilic and nonthermophilic sources in terms of structural parameters inferred from amino acid composition. Int. J. Peptide Protein Res. 10:39–50, 1977.
43. Bull, H.B., Breese, K. Thermal stability of proteins. Arch. Biochem. Biophys. 158:681–686, 1973.
44. Ikai, A. Thermostability and aliphatic index in globular proteins. J. Biochem. (Tokyo) 88:1895–1898, 1980.
45. Bigelow, C.C. On the average hydrophobicity of proteins and the relation between it and protein structure. J. Theor. Biol. 16:187–211, 1967.
46. Merkler, D.J., Farrington, G.K., Wedler, F.C. (1981). Protein thermostability. Int. J. Peptide Protein Res. 18:430–442, 1981.
47. Ponnuswamy, P.K., Muthusamy, R., Manavalan, P. Amino acid composition and thermal stability of proteins. Int. J. Biol. Macromol. 4:186–191, 1982.
48. Argos, P., Rossmann, M.G., Grau, U.M., Zuber, H., Frank, G., Tratschin, J.D. Thermal stability of proteins. Biochemistry 18:5698–5703, 1979.
49. Liao, H., McKenzie, T., Hageman, R. Isolation of a thermostable enzyme variant by cloning and selection in a thermophile. Proc. Natl. Acad. Sci. U.S.A. 83:576–580, 1986.
50. Yutani, K., Ogasahara, K., Tsujita, T., Sugino, Y. Dependence of conformational stability on hydrophobicity of the amino acid residue in a series of variant proteins substituted at a unique position of tryptophan synthase alpha subunit. Proc. Natl. Acad. Sci. U.S.A. 84:4441–4444, 1987.
51. Shindyalov, I.N., Kolchanov, N.A. Analysis of the factors and implications of an empirical method for estimating the stability of mutant human haemoglobins. J. Theor. Biol. 117:19–46, 1985.
52. Pantoliano, M.W., Ladner, R.C., Bryan, P.N., Rollence, M.L., Wood, J.F., Gilliland, G.L., Stewart, D.B., Poulos, T.L. The engineering of disulphide bonds, electrostatic interactions and hydrophobic contacts for the stabilization of subtilisin BPN'. Protein Engin. 1:229, 1987.
53. Vasantha, N., Thompson, L.D., Rhodes, C., Banner, D., Nagle, J., Filpula, D. Genes for alkaline protease and neutral protease from *Bacillus amyloliquiefaciens* containing a large open reading frame between the regions coding for the signal sequence and mature protein. J. Bacteriol. 159: 811–189, 1984.
54. Wells, J.A., Ferrari, E., Henner, D.J., Estell, D.A., Chen, E.Y. (1983). Cloning, sequencing and secretion of *Bacillus amyloliquefaciens* subtilisin in *Bacillus subtilis*. Nucl. Acid Res. 11:7911–7925, 1983.
55. Smith, E.H., Markland, F.S., Kasper, C.B., Delange, R.J., Landon, M., Evans, W.H. The complete amino acid sequence of two types of subtilisin, BPN' and Carlsberg. J. Biol. Chem. 241:5974–5976, 1966.
56. Jany, K.-D., Lederer, G., Mayer, B. Proteinase K—a new subclass of the subtilisins: The amino acid sequence of the enzyme. Biol. Chem. Hoppe-Seyler 367:87–87, 1986.
57. Kabsch, W., Sander, C. Dictionary of protein secondary structure. Pattern recognition of hydrogen bonded and geometrical features. Biopolymers 22:2577–2637, 1983.
58. Kabsch, W. A discussion of the solution for the best rotation to relate two sets of vectors. Acta Cryst. A 34:827–828, 1978.
59. Jones, A.T. A graphics model building and refinement system for macromolecules. J. Appl. Cryst. 11:268–272, 1978.
60. Sander, C. Physical criteria for folding units of globular proteins. In: "Structural Aspects of Recognition and Assembly in Biological Macromolecules" (M. Balaban et al., eds.). Rehovot and Philadelphia: Balaban ISS, 1981:183–195.
61. Martin, R.B. Bioinorganic chemistry of calcium. In: "Metal Ions in Biological systems" (H. Sigel, ed.), Vol 17. New York: Marcel Dekker, 1984:1–49.