Directed evolution converts subtilisin E into a functional equivalent of thermitase

Huimin Zhao[1] and Frances H.Arnold[2]

Division of Chemistry and Chemical Engineering 210–41, California Institute of Technology, Pasadena, CA 91125, USA

[1]Present address: Materials Research and Development, The Dow Chemical Company, 1707 Building, Midland, MI 48674, USA

[2]To whom correspondence should be addressed. E-mail: frances@cheme.caltech.edu

We used directed evolution to convert *Bacillus subtilis* subtilisin E into an enzyme functionally equivalent to its thermophilic homolog thermitase from *Thermoactinomyces vulgaris*. Five generations of random mutagenesis, recombination and screening created subtilisin E 5-3H5, whose half-life at 83°C (3.5 min) and temperature optimum for activity (T_{opt}, 76°C) are identical with those of thermitase. The T_{opt} of the evolved enzyme is 17°C higher and its half-life at 65°C is >200 times that of wild-type subtilisin E. In addition, 5-3H5 is more active towards the hydrolysis of succinyl-Ala-Ala-Pro-Phe-*p*-nitroanilide than wild-type at all temperatures from 10 to 90°C. Thermitase differs from subtilisin E at 157 amino acid positions. However, only eight amino acid substitutions were sufficient to convert subtilisin E into an enzyme equally thermostable. The eight substitutions, which include known stabilizing mutations (N218S, N76D) and also several not previously reported, are distributed over the surface of the enzyme. Only two (N218S, N181D) are found in thermitase. Directed evolution provides a powerful tool to unveil mechanisms of thermal adaptation and is an effective and efficient approach to increasing thermostability without compromising enzyme activity.

Keywords: *in vitro* evolution/StEP recombination/subtilisin E/thermitase/thermostability

Introduction

Understanding the molecular basis of protein thermostability is central to the dual goals of elucidating the relationship between protein sequence and structure and engineering improved proteins. Homologous enzymes can be found in organisms that grow in different environments, covering a temperature span of nearly 200°C. Comparisons of the structures and sequences of these evolutionarily related proteins, however, have yielded no general rules for thermostabilization. One reason for this is that stability reflects a complex interplay of forces; proteins are only marginally stable as a result of the delicate balance of numerous stabilizing and destabilizing interactions (Jaenicke *et al.*, 1996; Vogt and Argos, 1997; Richards, 1997). The other important reason is that homologous sequences generally differ at a substantial fraction of their amino acid positions, rendering it extremely difficult to identify mutations that confer changes in thermostability. Furthermore, if changes in thermostability are merely the products of random drift or of natural selection for other properties such as enzyme

activity (Giver *et al.*, 1998), it will be virtually impossible to obtain such rules by comparing natural enzymes (Benner and Ellington, 1988).

A more promising approach to elucidating molecular mechanisms of thermostabilization is to study engineered proteins in which the effects of specific amino acid substitutions can be pinpointed (Fersht and Serrano, 1993; Matthews, 1995). Specific mechanisms identified through such efforts, however, can prove difficult to implement for stabilization of useful proteins in a rational protein design strategy, particularly when few structural data are available. This is especially true if improvements in stability are not to come at the cost of some other property, such as enzyme activity. Engineering highly stable and active enzymes poses a difficult challenge to rational protein design.

In contrast, specific functional changes in proteins can be engineered efficiently by directed evolution. Inspired by natural Darwinian evolution, directed evolution involves mutation, recombination and screening or selection to accumulate the mutations required to achieve significant changes in protein function (Arnold, 1998). A wide variety of enzyme functions have been altered by this approach, including expression levels, enantioselectivity, substrate specificity and activity in non-natural environments (Kuchner and Arnold, 1997). Directed evolution can also be used to probe the molecular mechanisms underlying those improvements. Unlike natural evolution, directed evolution yields primarily adaptive changes in response to a specific set of 'selection pressures' defined by the experiment. Hence there is no large and confounding background of neutral mutations.

We are using directed evolution to probe the molecular basis of enzyme thermostability and to investigate the relationship between stability and catalytic activity (Giver *et al.*, 1998). Here we report the directed evolution of the alkaline serine protease subtilisin E from *Bacillus subtilis* to create a thermostable counterpart. Similarly to subtilisin BPN′ (with which it shares 80% sequence identity), subtilisin E is produced from a pre-pro-subtilisin consisting of a 29-residue pre-sequence, a pro-sequence of 77 residues and the mature protease of 275 residues (Stahl and Ferrari, 1984). The pre-sequence functions as the signal peptide for protein secretion across the membrane (Wong and Doi, 1986), while the pro-sequence acts as a 'foldase' to guide the folding of the subtilisin molecule (Ohta and Inouye, 1990; Gallagher, *et al.*, 1995). Subtilisin E shares only 43% amino acid identity (157 amino acid differences) with its thermophilic homolog thermitase from *Thermoactinomyces vulgaris* (Frömmel *et al.*, 1978; Meloun *et al.*, 1985). Subtilisin E inactivates within minutes at 65°C, whereas thermitase requires temperatures above 80°C for rapid inactivation (Schreier *et al.*, 1984). We have used directed evolution to convert subtilisin E into a functional equivalent to thermitase.

Materials and methods

Restriction enzymes were obtained from Boehringer Mannheim (Indianapolis, IN) and succinyl-Ala-Ala-Pro-Phe-*p*-nitroanilide

(s-AAPF-*p*Na) from Sigma (St Louis, MO). Purified thermitase was kindly provided by Dr Wolfgang Höhne (Humboldt University, Germany).

Random mutagenesis, recombination and thermostability screening

Random mutagenesis was carried out by error-prone PCR and *in vitro* recombination by the staggered extension process (StEP) as described (Zhao *et al.*, 1998). Purified restricted inserts from PCR and StEP reactions were ligated with vectors generated by *Bam*HI–*Nde*I digestion of pBE3 *Escherichia coli–B. subtilis* shuttle vector. *E.coli* HB101 were transformed by electroporation and selected with 100 μg/ml ampicillin. A typical ligation reaction (10 μl final volume) contained 1× T4 DNA ligase buffer, ~12 ng inserts, ~50 ng vector and 0.5 U T4 DNA ligase and was incubated at 15–16°C for 16 h. The resulting plasmid library was isolated and transformed into *B.subtilis* DB428. Colonies were grown in 96-well plates containing 200 μl Schaeffer's (SG) medium with 30 μg/ml kanamycin. After 20 h at 37°C, the plates were centrifuged at 3000 r.p.m. for 10 min and 5 μl of supernatant were transferred into two fresh plates. A 15 μl volume of SG medium was added. Initial activity was measured after 5 min of incubation in an oven set at a specific temperature; residual activity was measured after 20 min. Machined aluminum blocks were used for uniform heating and the 5 min incubation for initial activity measurement reduces variability in the assay. Enzyme activity was assayed at 37°C after addition of 100 μl of 0.2 mM s-AAPF-*p*Na in assay solution (10 mM Tris–HCl, 1 mM $CaCl_2$, pH 8.0, prewarmed at 37°C). The activity was determined from the absorbance at 405 nm within 1 min in a ThermoMax 96-well plate reader (Molecular Devices).

Enzyme kinetics

Subtilisins were purified as described (Zhao and Arnold, 1997a). All kinetic measurements were made using a thermo-stated Milton-Roy Spectronic 3000 spectrophotometer. Purified enzymes were dialyzed in assay solution at 4°C overnight before characterization. Enzyme concentration was determined by measuring A_{280} (subtilisin E, $\varepsilon = 35\,886$ M^{-1} cm^{-1}; thermitase, $\varepsilon = 59\,055$ M^{-1} cm^{-1}). Initial rates of hydrolysis of s-AAPF-*p*Na in 0.99 ml of assay solution at 37°C and specific activities were determined as described (Zhao *et al.*, 1998). To determine k_{cat} and K_M for each subtilisin, initial rates were determined at eight different substrate concentrations over the range 0.02–1.5 mM that bracketed K_M. For thermitase, the substrate concentration was 0.002–0.08 mM. Data were fitted to the Michaelis–Menten equation by non-linear regression. T_{opt} was determined by incubating enzymes with s-AAPF-*p*Na (2 mg/ml) in assay solution at different temperatures for 10 min, after which the product *p*Na was determined by measuring A_{410}.

Half-lives of thermal inactivation

Incubations were carried out in an MJ Research (Watertown, MA) PTC-200 thermocycler for precise temperature control. Purified enzymes were incubated in assay solution at specific temperatures. Aliquots taken at various time intervals were diluted into 1.0 ml of assay solution with 0.2 mM s-AAPF-*p*Na at 37°C. Typically, inactivation was followed until >80% of the activity was lost. Plots of log(residual activity) versus time were linear. Inactivation rate constants (k_{inact}) were obtained from the slope and half-lives were calculated as $t_{\frac{1}{2}} = ln2/k_{inact}$.
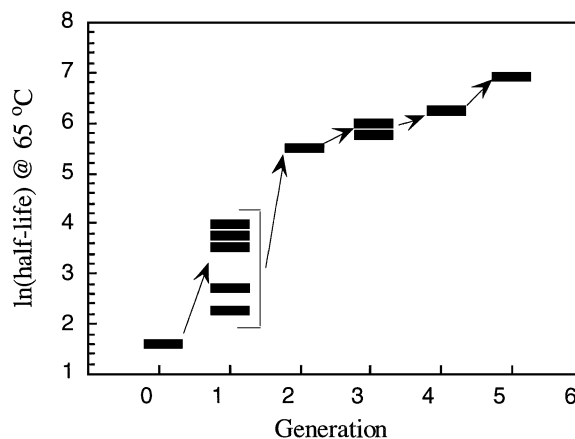
Fig. 1. Evolutionary strategy and progression of subtilisin E thermostability, as measured by ln(half-life) at 65°C in 10 mM Tris–HCl (pH 8.0), 1 mM $CaCl_2$. Random mutagenesis and screening yielded five thermostable first-generation variants, which were recombined to create the second-generation library, from which the single most thermostable variant was selected. A second round of mutagenesis and screening yielded three thermostable third-generation variants, which were recombined and screened to identify the most stable fourth-generation mutant. A final round of random mutagenesis and screening yielded 5-3H5.

T_{50}

Values of T_{50}, the temperature at which 50% of the protease activity is lost during a fixed incubation period, were determined as described (Eijsink *et al.*, 1992). Aliquots of purified subtilisins were incubated for 20 min at various temperatures. Residual activity was determined and expressed as a percentage of the initial activity measured at 37°C.

Results

Experimental strategy and directed evolution of subtilisin E

A key element of any directed evolution experiment is the development of a rapid screen sensitive enough to ensure that the expected small changes in thermostability brought about by single amino acid substitutions can be observed (Zhao and Arnold, 1997b). The screen used here is based on the retention of activity after incubation at an elevated temperature (Bryan *et al.*, 1986) and is not designed to distinguish the various mechanisms of enzyme inactivation. Variants whose ratio of residual to initial activity indicated greater thermostability than the parent and whose initial activity was comparable to that of wild-type at 37°C were selected for further analysis. Thermostabilization was verified by measuring the half-life of thermal inactivation of the purified enzyme. The incubation temperature for screening, chosen as the point where the residual activity of the parent after incubation is 30–40% of its initial activity, was gradually elevated with each generation of directed evolution.

Error-prone PCR was used to create a subtilisin E gene library with an average of 2–3 base changes per gene, a rate that maximizes the population of single amino acid mutants and greatly simplifies analysis of the products. Approximately 5000 clones were screened for activity and thermostability, with incubation at 65°C. Of the positives, five were confirmed to have longer half-lives at 65°C than wild-type (Figure 1). Equal amounts of these five genes were then recombined using StEP in order to identify the best combination of beneficial mutations present in the parent genes (and at the same time remove any deleterious mutations). Approximately 8000 clones

Table I. Thermostability of wild-type and evolved subtilisins E

| Generation	Subtilisin E variant	dT_{50} (°C)	$t_{1/2}$ (min)			
			65°C	75°C	80°C	83°C
0	Wild-type	0	4.9			
	4A5	8.2	44.2			
	15C1	7.6	35.5			
1	32G11	2.3	9.8			
	35F10	7.8	39.5			
	36D10	3.1	15.2			
2	45B5	13.1	250	8.7		
	5H2	14.0	345	12.0		
3	16D11	14.1	368	12.8		
	20E8	14.2	391	13.6		
4	8B3	15.2	526	18.3	5.3	
5	3H5	17.2	1030		10.3	3.53

Values of $t_{1/2}$ at each temperature are the average of three experiments (standard deviation <5%). dT_{50} is the increase in T_{50} in the evolved thermostable variants with respect to wild-type (59.2°C). All experiments were performed in 10 mM Tris–HCl (pH 8.0), 1 mM CaCl$_2$. Mutations in each variant are listed in Figure 4.

from the recombined library were screened, with incubation at 75°C. Clone 2-45B7 exhibited the highest thermostability index and the longest half-life at 75°C. This variant was used to parent a second round of random mutagenesis. From this library, ~2000 clones were screened for thermostability at 76°C. Three variants were identified and confirmed to have longer half-lives than 2-45B7. These three genes were recombined again by StEP and ~3000 clones were screened at 78°C to produce variant 4-8B3 with the highest thermostability index and the longest half-life. A third round of mutagenesis and screening ~2000 clones at 80°C yielded subtilisin E 5-3H5. The overall strategy and progression towards improved thermostability are summarized in Figure 1.

Thermostabilities of evolved subtilisins

At elevated temperatures, subtilisin inactivates irreversibly through autolysis (Wells and Powers, 1986; Mitchinson and Wells, 1989; Vriend and Eijsink, 1993). Preventing autolysis during stability measurements is extremely difficult and a rigorous thermodynamic analysis of subtilisin stability is not possible (Mitchinson and Wells, 1989; Vriend and Eijsink, 1993). Therefore, subtilisin thermostability is usually measured by the kinetics of inactivation, reported either as the rate of thermal inactivation (or half-life, $t_{1/2}$) (Mitchinson and Wells, 1989) or T_{50}, the temperature at which 50% is lost during a fixed incubation period (Eijsink *et al.*, 1991; Vriend and Eijsink, 1993).

Table I lists the $t_{1/2}$ and T_{50} values of wild-type and the 11 evolved thermostable subtilisins E collected during the five generations of directed evolution. At 65°C, the $t_{1/2}$ of wild-type subtilisin E is ~5 min and its T_{50} is 59.2°C. The first generation variants have half-lives 2–8 times longer than wild-type at 65°C and their T_{50}s are 2.3–8.2°C higher. The half-life of their recombined product 2-45B7 is ~50 times that of wild-type and its T_{50} is increased by 13.1°C. A further round of point mutagenesis in the third generation prolongs the half-lives by ~50% at 75°C and increases the T_{50}s by ~1°C. The recombined product 4-8B3 shows a further increase in $t_{1/2}$ at 75°C; its T_{50} of 4-8B3 is now 15.2°C above that of wild-type. The half-life of the fifth generation enzyme 5-3H5 is twice that of 4-8B3 at 80°C and at least 200 times that of wild-type
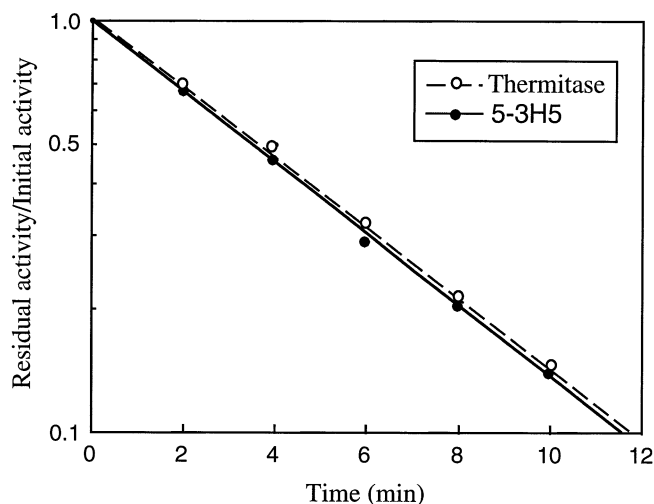
Fig. 2. Kinetics of thermal inactivation of subtilisin 5-3H5 and thermitase at 83°C in 10 mM Tris–HCl (pH 8.0), 1 mM CaCl$_2$.

Table II. Specific activities and kinetic constants of wild-type and evolved subtilisin E variants, in 10 mM Tris–HCl (pH 8.0), 1 mM CaCl$_2$ at 37°C

Generation	Subtilisin E variant	Specfic activity (U/mg)	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M (s^{-1} mM^{-1})
0	Wild-type	20.0	25.4	0.385	66.0
	4A5	40.0	52.4	0.388	135
	15C1	21.0	29.4	0.384	76.6
1	32G11	44.2	30.6	0.158	193
	35F10	40.6	52.3	0.383	136
	36D10	19.8	25.6	0.385	66.5
2	45B7	69.9	55.6	0.152	366
	5H2	71.6	58.8	0.155	379
3	16D11	73.0	57.1	0.150	381
	20E8	72.2	56.2	0.151	372
4	8B3	72.5	56.5	0.153	369
5	3H5	71.0	55.8	0.151	373
	Thermitase	293	130	0.0146	8910

Values of specific activities are the average of five measurements (standard deviation <5%). Mutations in each variant are listed in Figure 4.

at 65°C (Figure 1). The T_{50} of 5-3H5 is 76.4°C, an increase of 17.2°C compared with wild-type.

Irreversible inactivation of subtilisin E at 65°C follows first-order kinetics and the $t_{1/2}$ is independent of enzyme concentration (data not shown). The inactivation kinetics of the evolved thermostable subtilisin E variants are also first order and independent of enzyme concentration. Thermitase inactivation at 83°C is also first order (Figure 2) and independent of enzyme concentration (Zhao, 1998).

Activities of evolved subtilisins

Specific activities of the wild-type and evolved thermostable subtilisin variants towards the peptide substrate s-AAPF-pNa are listed in Table II. First-generation variants 1-4A5, 1-32G11 and 1-35F10 are slightly more active than wild-type; 1-4A5 and 1-35F10 have k_{cat} twice that of wild-type, while their K_M remain unchanged; 1-32G11 exhibits about a 2-fold decrease in K_M and a slight increase in k_{cat}. All variants in subsequent generations have similar k_{cat} and K_M. The catalytic efficiencies of the generation 2–5 subtilisins are 5–6 times greater than that of wild-type.
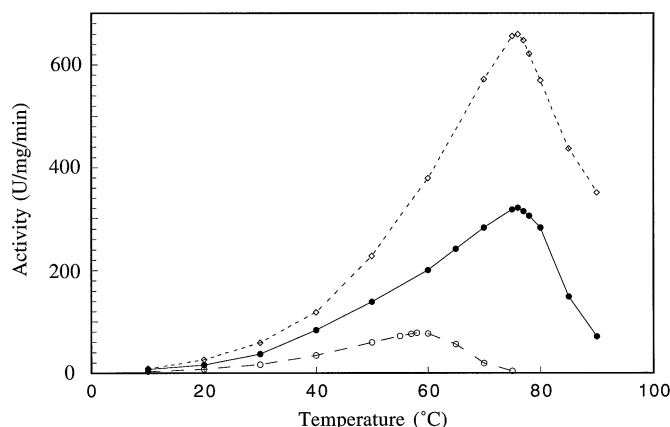
Fig. 3. Activity–temperature profiles of wild-type subtilisin E (– – –), 5-3H5 (—) and thermitase (- - -).

Comparison with thermitase

Inactivation half-lives for thermitase and subtilisin 5-3H5 were measured at 83°C. As shown in Figure 2, both enzymes exhibit first-order inactivation kinetics with a $t_{1/2}$ of 3.5 min. This is similar to the reported thermitase $t_{1/2}$ of 1.2 min (83°C in 0.1 M HEPES buffer, pH 7.6, 1 mM Ca^{2+} (Schreier *et al.*, 1984)). Activity–temperature profiles of wild-type, 5-3H5 subtilisin and thermitase are shown in Figure 3. Both 5-3H5 and thermitase have temperature optima at ~76°C, which is 18°C higher than that of wild-type subtilisin E, which is in good agreement with the observed increases in T_{50} (Table I). The activity–temperature profile of 5-3H5 is significantly broadened with respect to the wild-type enzyme and is elevated at all temperatures. As shown in Table II, the specific activity of thermitase at 37°C is ~15 times that of subtilisin E. The higher catalytic efficiency is mostly a product of its lower K_M for this peptide substrate.

Sequence analysis

DNA sequences of the 11 thermostable subtilisin E genes were determined in an effort to probe the molecular basis of thermostabilization and the evolutionary pathway by which the enzyme was stabilized (Figure 4). All together these sequences contain 21 base substitutions, 11 of which cause amino acid changes. The 21 DNA substitutions are distributed throughout the targeted mature subtilisin E gene. Two base changes never appeared simultaneously in a single codon and the types of substitutions are more or less balanced with regard to transitions (10) and transversions (11).

Amino acid substitutions S161C, G166R, S194P, N181D and N218S have been identified as thermostable mutations in the first-generation variants, while S37T and A192T are neutral (Zhao *et al.*, 1998). Third-generation mutations S9F, P14L and N76D are thermostabilizing since they are the only non-synonymous mutations in their respective genes. P14L and N76D are recombined in the fourth-generation variant 4-8B3. The failure to identify a variant with known beneficial mutation S9F can occur for one of two reasons (Zhao *et al.*, 1998): (1) S9F does not contribute to thermostability in the background of the other mutations or (2) S9F is too close to P14L (only 15 bp apart) to recombine, and therefore recombinant thermostable genes will contain one or the other, but not both. N118S increases the thermostability of wild-type subtilisin E (Chu *et al.*, 1995) and is therefore believed to be a stabilizing mutation in 5-3H5. It is noteworthy that S161C appeared twice, in two different generations.
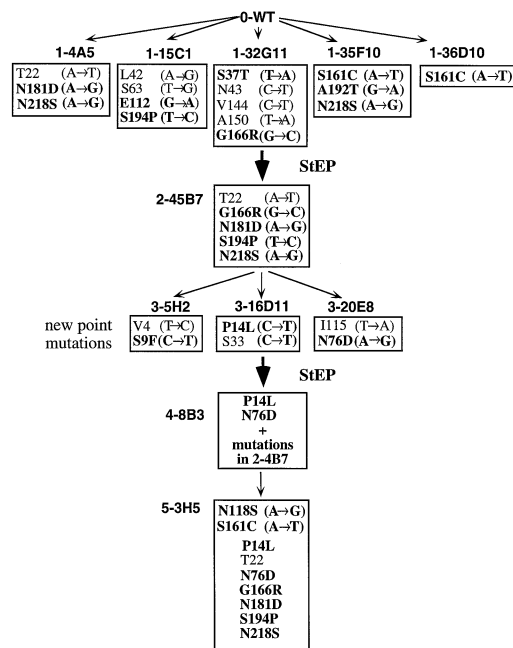

Fig. 4. Lineage and mutations in evolved thermostable subtilisin E variants. Base mutations are shown in parentheses.

Table III. Effects of amino acid substitutions on thermostability and activity (hydrolysis of s-AAPF-*p*NA) of subtilisin E

	Stability	Activity
S9F	+	0
P14L	+	0
S37T	0	0
N76D	+	0
N118S	+	0
S161C	+	0
G166R	+	+
N181D	+	0
A192T	0	0
S194P	+	(+)
N218S	+	+

The effects of individual mutations on catalytic activity can also be discerned. For example, S161C does not affect activity because the k_{cat} and K_M of 1-36D10, which only contains the S161C substitutution, are similar to those of wild-type (Table II, Figure 4). S194P (1-15C1) has a small effect. We found previously that N181D has no effect on specific activity, whereas N218S increases it 2-fold (Zhao and Arnold, 1997a). These mutations are additive in 1-4A5 (N181D + N218S), which has the same K_M as wild-type but a 2-fold larger k_{cat}. Effects of the 11 amino acid substitutions on stability and catalytic activity are summarized in Table III.

Discussion

Evolution strategy

This experiment implemented an evolutionary strategy consisting of random mutagenesis and screening followed by recombination of a few of the best mutants. A round of random mutagenesis and screening 1000–2000 clones typically yields at least several true thermostable variants. Using only one to parent another generation of random mutagenesis would waste the information contained in the other thermostable genes.

Hence the most thermostable variants were recombined and the gene containing the best combination of mutations was used to parent the next round of random mutagenesis (experimental and mutation results are summarized in Figure 4). An alternative might be to use recursive cycles of recombination, relying on the point mutagenesis inherent in the recombination operation to create new mutations. Implementing recombination and point mutagenesis in the same operation, however, is likely to lead to suboptimal results. Recombination increases screening requirements, because a large fraction of the sequences created (75%) are parents and grandparents (Moore *et al.*, 1997). While recombination has the potential to create novel combinations of beneficial mutations by combining two neutral or even deleterious ones (non-additive effects), there is little evidence that this plays an important role in the evolution of single sequences. Because screening requirements increase rapidly with increasing number of parent sequences, we recombined only the most thermostable genes (<6) so that all possible combinations could be explored. Simple recombination statistics and alternative search strategies were discussed by Moore *et al.* (1997).

Directed evolution has increased the $t_{1/2}$ of subtilisin E at 65°C more than 200-fold. The evolved enzyme has a $t_{1/2}$ at 83°C comparable to its thermophilic homolog, thermitase. Similarly to what we observed during the directed evolution of a thermostable esterase (Giver *et al.*, 1998), we find that enhancing thermostability while maintaining catalytic activity at lower temperature (37°C) is sufficient to create an enzyme that is highly active at elevated temperatures. The activities of the evolved subtilisins E increase with increasing temperature until the enzymes denature. The increase in the temperature of optimum activity T_{opt} mirrors the increase in T_{50} (and $t_{1/2}$). Hence it is not necessary to screen directly for activity at high temperature in order to evolve a 'thermophilic' enzyme.

Sequence comparisons

Subtilisin E belongs to the superfamily of subtilisin-like serine proteases ('subtilases'), of which more than 200 members are known (Siezen and Leunissen, 1997). Multiple sequence alignments indicate a high degree of sequence flexibility within this family. With the exception of the essential catalytic triad residues (D32, H64 and S221 in subtilisin E) and a single glycine residue (G219), it appears that virtually every other residue has been replaced by one or more amino acids during evolution. Large deletions and insertions are also frequently found. Subtilisin E homologs from thermophiles or hyperthermophiles have also been characterized, including thermitase from *Thermoactinomyces vulgaris* (Frömmel *et al.*, 1978; Meloun *et al.*, 1985), aqualysin I from *Thermus aquaticus* YT-1 (Matsuzawa *et al.*, 1988) and aerolysin from *Pyrobaculum aerophilum* (Volkl *et al.*, 1994). Thermitase, for example, differs from subtilisin E at 157 amino acid positions. However, eight stabilizing mutations are sufficient to convert subtilisin E into an enzyme of equivalent thermostability. Of the eight stabilizing mutations in 5-3H5 subtilisin E, only N218S and N181D are found in thermitase. As in 5-3H5, however, residue 166 in thermitase also has a long side chain (N).

Based on the subtilisin sequences aligned by Siezen and Leunissen (1997), all 11 amino acid substitutions that occurred during directed evolution of subtilisin E are found in the sequences of at least one other naturally occurring subtilisin. Hence these amino acid substitutions have been tried in nature and found to be acceptable. However, it is impossible to
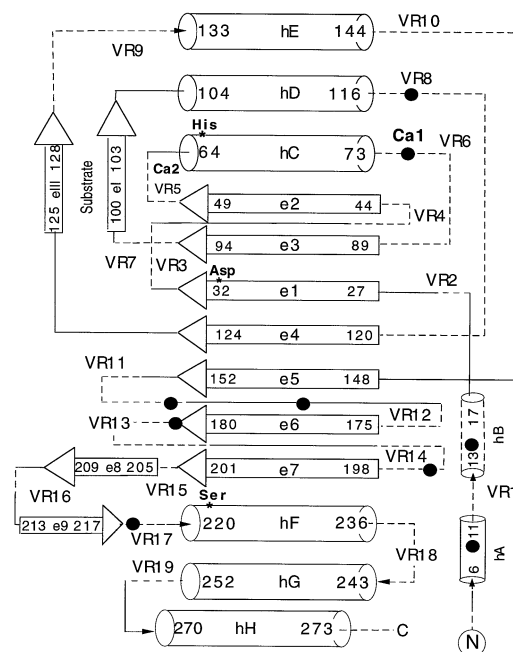

Fig. 5. Schematic representation of the secondary structure and topology of subtilisin E, with α-helices shown as cylinders and β-sheet strands as arrows. Solid lines indicate the conserved regions and dashed lines the variable regions. Approximate locations of the main Ca^{2+}-binding sites (Ca1 and Ca2), catalytic triad residues D32, H64, S221 (by asterisks) and substrate-binding region (between strands eI and eIII) are indicated. (*) Locations of stabilizing amino acid substitutions. The most thermostable variant 5-3H5 contains all these mutations except for S9F. Drawn after Siezen and Leunissen (1997).

identify stabilizing mutations based on sequence comparisons of subtilisins from mesophilic and (hyper-)thermophilic organisms. For example, thermostabilizing mutation S161C is found only once in a mesophilic subtilase from *Drosophila melanogaster*. G166R appears in another from *Bacillus subtilis*. Neither appears in the known natural thermostable subtilisins. Similarly, thermostabilizing mutations S9F and P14L are found exclusively in mesophilic subtilases. The well known thermostabilizing mutation N218S (Bryan *et al.*, 1986) is found in 67 out of 127 aligned sequences, including psychrophilic, mesophilic, thermophilic and hyperthermophilic subtilisins. On the other hand, neutral mutation S37T is also found in thermophilic subtilisins.

It has been suggested that certain amino acid residues or certain amino acid replacement pairs are used preferentially in thermophilic proteins (Singleton *et al.*, 1977; Ikai, 1980; Ponnuswamy *et al.*, 1982; Menendez-Arias and Argos, 1989). For example, Lys→Arg and Ser→Ala are more commonly found than other substitutions (Menendez-Arias and Argos, 1989). However, none of these replacements was needed to convert subtilisin E into its thermostable counterpart. Nor did we observe other proposed thermo-adaptive changes, such as increases in alanine, isoleucine or tyrosine content (Russell and Taylor, 1995). It is likely that each mutation will need to be evaluated in its specific structural context.

Structural analysis

Two subtilisin structural models were generated to examine further the molecular basis of thermostability and activity for the stabilizing mutations. The first, shown in Figure 5, maps these mutations on to a cartoon of the secondary structure and topology (Siezen and Leunissen, 1997). Most of the
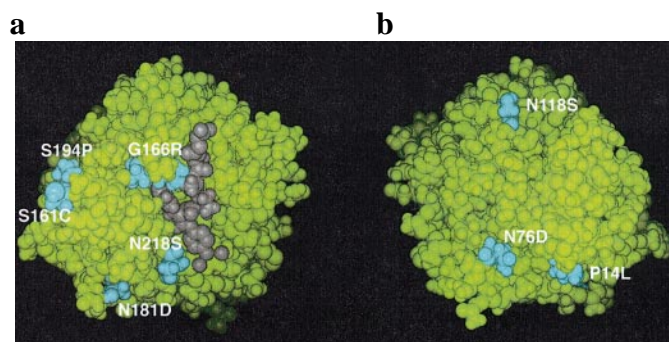
Fig. 6. Space-filling model of 5-3H5 subtilisin E showing the eight thermostabilizing mutations (cyan) and peptide substrate s-AAPF-*p*Na (gray). (**b**) View after rotation of (**a**) by 180°.

thermostabilizing mutations are located in loops connecting elements of secondary structure. The substitution of Ser194 with Pro probably reduces the entropy of the flexible loop. Only two mutations (S9F and P14L) are located in helices. The proline to leucine substitution (P14L) in helix-B may increase stability by increasing helical propensity. Only two mutations, G166R and S9F, occur in evolutionarily-conserved regions (solid lines). Most of the stabilizing mutations appear in variable regions (dashed lines), as was also found for mutations that enhance subtilisin activity in non-aqueous solvents (Chen and Arnold, 1993).

Figure 6 shows the positions of the thermostabilizing amino acid substitutions in a three-dimensional structure model generated using the known structure of wild-type subtilisin E (Chu *et al.*, 1995). The amino acid substitutions are all distributed over the surface of the enzyme (G166R is partially exposed) and most are far from the active site. Residues 166 and 218, however, are involved in substrate binding, which explains their effects on catalytic activity. Residue 166 is located at the bottom end of the distinct, elongated cleft known as S1 pocket, which determines the substrate specificity. Residue 218 is involved in the binding of the C-terminal or leaving portion (P1′–P2′) of the substrate. The stabilizing effect of G166R may be attributed to the reduced cavity volume. N218S presumably increases thermostability by slight improvement of hydrogen bonding near a β-bulge, the thermostabilizing mechanism found in subtilisin BPN′ (Bryan *et al.*, 1986). N76D is believed to stabilize the enzyme by enhancing calcium ion binding (Pantoliano *et al.*, 1989). The mechanisms by which S9F, S161C, N181D and N118S enhance thermostability are currently under investigation.

The fact that all the stabilizing mutations are located on the protein surface is consistent with a model in which subtilisin thermal inactivation is largely determined by local unfolding processes. Since the early steps of unfolding of a protein are thought to involve mainly surface-located structure elements (Matouschek *et al.*, 1989; Jackson and Fersht, 1991), mutations stabilizing these elements should enhance enzyme stability. However, the apparent additivity of at least several of the mutations argues against stabilization based on mere interference with local unfolding. Single surface mutations are also more easily accommodated in the protein structure. Buried mutations are much less likely to enhance stability without compensating changes elsewhere. We would expect that specific sets of internal mutations will also stabilize this enzyme. However, an evolutionary process involving primarily

single amino acid substitutions is unable to uncover these rare events (Giver *et al.*, 1998).

Conclusions

Much effort has been devoted to understanding protein stability and generating thermostable enzymes. Improved thermal stability is useful for applications involving exposure to high temperatures; it is also useful in that it is often accompanied by increased resistance to inactivation by denaturants (e.g. by organic solvents) (Cowan, 1997) and proteolysis. Future work to characterize the stabilities of the evolved subtilisins under different conditions will help us to assess how well a simple screen for thermostability will extend to evolving overall enzyme stability.

A relatively small number of amino acid substitutions is needed to convert a mesophilic enzyme into a variant as stable as its thermophilic counterpart, without sacrificing catalytic activity. This experiment rediscovered known subtilisin-stabilizing mutations and also a number of others not previously described. The stabilizing mutations are distributed over the protein surface and are structurally isolated. Their effects on thermostability are presumed to be independent and, to a large extent, additive. Various stabilization mechanisms have been recruited. The present study strongly supports the notion that enzymes are stabilized by the cumulative effects of small improvements at many locations within the protein molecule (Jaenicke *et al.*, 1996). Thus the 'holy grail' for protein stabilization, rules that are general and easy to apply, probably does not exist. Nevertheless, directed evolution is a 'holy approach', a design algorithm that requires no structural information or knowledge of the principles governing protein stability. When the selective pressure includes all properties of interest ('you get what you screen for'), useful enzymes can be obtained.

References

Arnold,F.H. (1998) *Acc. Chem. Res.*, **31**, 125–131.
Benner,S. and Ellington,A.D. (1988) *CRC Crit. Rev. Biochem.*, **23**, 369–426.
Bryan,P.N., Rollence,M.L., Pantoliano,M.W., Wood,J., Finzel,B.C., Gilliland,G.L., Howard,A.J. and Poulos,T.L. (1986) *Proteins: Struct. Funct. Genet.*, **1**, 326–334.
Chen,K. and Arnold,F.H. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 5618–5622.
Chu,N.M., Chao,Y. and Bi,R.C. (1995) *Protein Engng*, **8**, 211–215.
Cowan,D.A. (1997) *Comp. Biochem. Physiol.*, **118A**, 429–438.
Eijsink,V., Van den Burg,B., Vriend,G., Berendsen,H. and Venema,G. (1991) *Biochem. Int.*, **24**, 517–525.
Eijsink,V.G., Vriend,G., van der Vinne,B., Hazes,B., van den Burg,B. and Venema,G. (1992) *Proteins: Struct. Funct. Genet.*, **14**, 224–236.
Fersht,A.R. and Serrano,L. (1993) *Curr. Opin. Struct. Biol.*, **3**, 75–83.
Frömmel,C., Hausdorf,G., Wöhne,W.E., Behnke,U. and Ruttloff,H. (1978) *Acta Biol. Med. Ger.*, **37**, 1193–1204.
Gallagher,T., Gilliland,G. and Bryan,P. (1995) *Structure*, **3**, 907–914.
Giver,L., Gershenson,A., Freskgard,P.-O. and Arnold,F. H. (1998) *Proc. Natl Acad. Sci. USA*, **95**, 12809–12813.
Ikai,A. (1980) *J. Biochem. (Tokyo)*, **88**, 1895–1898.
Jackson,S.E. and Fersht,A.R. (1991) *Biochemistry*, **30**, 10436–10443.
Jaenicke,H., Schurig,H., Beaucamp,N. and Ostendorp,R. (1996) *Adv. Protein Chem.*, **48**, 181–269.
Kuchner,O. and Arnold,F.H. (1997) *Trends Biotechnol.*, **15**, 523–530.
Matouschek,A., Kellis,J.T., Serrano,L. and Fersht,A.R. (1989) *Nature*, **340**, 122–126.

Matsuzawa,H., Tokugawa,K., Hamaoki,M., Mizoguchi,M., Taguchi,H., Terada,I., Kwon,S.T. and Ohta,T. (1988) *Eur. J. Biochem.*, **171**, 441–447.

Matthews,B.W. (1995) *Adv. Protein Chem.*, **46**, 249–278.

Meloun,B., Baudys,M., Kostka,V., Hausdorf,G., Frömmel,C. and Höhne,W.E. (1985) *FEBS Lett.*, **183**, 195–200.

Menendez-Arias,L. and Argos,P. (1989) *J. Mol. Biol.*, **206**, 397–406.

Mitchinson,C. and Wells,J.A. (1989) *Biochemistry*, **28**, 4807–4815.

Moore,J.C., Jin,H.M., Kuchner,O. and Arnold,F.H. (1997) *J. Mol. Biol.*, **272**, 336–347.

Ohta,Y. and Inouye,M. (1990) *Mol. Microbiol.*, **4**, 295-304.

Pantoliano,M.W., Whitlow,M., Wood,J.F., Dodd,S.W., Hardman,K.D., Rollence,M.L. and Bryan,P.N. (1989) *Biochemistry*, **28**, 7205–7213.

Ponnuswamy,P.K., Muthusamy,R. and Manavalan,P. (1982) *Int. J. Biol. Macromol.*, **4**, 186–191.

Richards,F.M. (1997) *Cell. Mol. Life Sci.*, **53**, 790–802.

Russell,R.J.M. and Taylor,G.L. (1995) *Curr. Opin. Biotechnol.*, **6**, 370–374.

Schreier,E., Fittkau,S. and Höhne,W.E. (1984) *Int. J. Pept. Protein Res.*, **23**, 134–141.

Siezen,R. and Leunissen,J.A.M. (1997) *Protein Engng*, **6**, 501–523.

Singleton,R., Middaugh,C.R. and McElroy,R.P. (1977) *Int. J. Pept. Protein Res.*, **10**, 39–50.

Stahl,M.L. and Ferrari,E. (1984) *J. Bacteriol.,* **158**, 411–418.

Vogt,G. and Argos,P. (1997) *Folding Des.*, **2**, S40–S46.

Volkl,P., Markiewicz,P., Stetter,K.O. and Miller,J.H. (1994) *Protein Sci.*, **3**, 1329–1340.

Vriend,G. and Eijsink,V. (1993) *J. Comput.-Aided Mol. Des.*, **7**, 367–396.

Wells,J.A. and Powers,D.B. (1986) *J. Biol. Chem.*, **261**, 6564–6570.

Wong,S.L. and Doi,R.H. (1986) *J. Biol. Chem.*, **261**, 10176–10181.

Zhao,H. (1998) PhD Thesis, California Institute of Technology.

Zhao,H. and Arnold,F.H. (1997a) *Proc. Natl Acad. Sci. USA*, **94**, 7997–8000.

Zhao,H. and Arnold,F.H. (1997b) *Curr. Opin. Struct. Biol.*, **7**, 480–485.

Zhao,H., Giver,L., Shao,Z., Affholter,J.A. and Arnold,F.H. (1998) *Nature Biotechnol.*, **16**, 258–262.