Protein Engineering vol.9 no.1 pp.77–83, 1994

Directed evolution of subtilisin E in *Bacillus subtilis* to enhance total activity in aqueous dimethylformamide

L.You[1] and F.H.Arnold[2]

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

[1]Present address: Bristol-Myers Squibb Co., Bio/Chem Division, PO Box 4744, Syracuse, NY 13221, USA

[2]To whom correspondence should be addressed

Sequential rounds of error-prone PCR to introduce random mutations and screening of the resultant mutant libraries have been used to enhance the total catalytic activity of subtilisin E significantly in a non-natural environment, aqueous dimethylformamide (DMF). Seven DNA substitutions coding for three new amino acid substitutions were identified in a mutant isolated after two additional generations of directed evolution carried out on 10M subtilisin E, previously 'evolved' to increase its specific activity in DMF. A *Bacillus subtilis–Escherichia coli* shuttle vector was developed in order to increase the size of the mutant library that could be established in *B.subtilis* and the stringency of the screening process was increased to reflect total as well as specific activity. This directed evolution approach has been extremely effective for improving enzyme activity in a non-natural environment: the resulting-evolved 13M subtilisin exhibits specific catalytic efficiency towards the hydrolysis of a peptide substrate succinyl-Ala-Ala-Pro-Phe-*p*-nitroanilide in 60% DMF solution that is three times that of the parent 10M and 471 times that of wild type subtilisin E. The total activity of the 13M culture supernatant is enhanced 16-fold over that of the parent 10M.

Keywords: *B.subtilis*/directed evolution/subtilisin E/total activity/random mutagenesis/organic solvents

Introduction

Available natural enzyme resources are being tailored to fulfil increasing demands for new biocatalysts. A practical strategy for altering enzyme properties is to introduce random base substitutions into the gene sequence and then select or screen for variants that express the desired phenotype(s). Features that have been enhanced by random mutagenesis include catalytic activity (Graham *et al.*, 1993), activity in organic solvents (Chen and Arnold, 1993), thermostability (Bryan *et al.*, 1986; Liao *et al.*, 1986; Joyet *et al.*, 1992), alkaline stability (Cunningham and Wells, 1987) and substrate specificity (Oliphant and Struhl, 1989; Graham *et al.*, 1993). Once the genes are sequenced, effective mutations can be accumulated by site-directed mutagenesis (Joyet *et al.*, 1992; Strausberg *et al.*, 1995). An attractive alternative to sequencing and site-directed mutagenesis is to accumulate beneficial mutations in sequential rounds of random mutagenesis (Chen and Arnold, 1993) or by a novel recombination approach (Stemmer, 1994), following a 'directed evolution' strategy (Arnold, in press). Directed evolution is likely to prove useful in enhancing enzyme performance in 'non-natural' environments (Arnold, 1993) as well as for obtaining new features never required by nature, provided an efficient selection or screening method can be found to channel the enzyme's evolution towards the desired properties. A significant advantage of this approach over 'rational' design methods is that neither structural information nor a mechanistic road-map are required to guide the directed evolution experiment.

In addition to creating new enzymes for applications in biotechnology, directed evolution methods can be used to explore the limits of protein function. Although naturally occurring enzymes have evolved within the context of living organisms, there is no reason to believe that they are limited to the relatively narrow set of conditions that will support life. Water-miscible organic solvents such as dimethylformamide (DMF), for example, are highly toxic to most organisms, even at levels of <10%. Although enzymes can often tolerate much higher concentrations and still retain their folded structures, catalytic activity is often significantly compromised. To probe whether enzymes could be 'tuned' to tolerate and even thrive in an apparently hostile environment (high concentrations of DMF), we applied a directed evolution strategy involving sequential generations of polymerase chain reaction (PCR) random mutagenesis and screening to the serine protease subtilisin E (Chen and Arnold, 1993). Mutations that improved the specific activity of subtilisin in aqueous DMF were not rare; mutations conferring ~2-fold enhancements in specific proteolytic activity could be found by screening only a few hundred colonies from each generation (Chen and Arnold, 1993). The accumulation of effective mutations in sequential rounds of random mutagenesis led to a significant enhancement of the enzyme's activity: the specific activity of the 'evolved' enzyme containing 10 amino acid substitutions (10M subtilisin E) was enhanced 157-fold in 60% DMF, recovering much of the activity lost by addition of the organic solvent.

Subtilisin is a useful catalyst for organic synthesis, particularly in the presence of organic solvents. Subtilisin can catalyze regioselective (Riva and Klibanov, 1988; Riva *et al.*, 1988) and stereoselective (Margolin *et al.*, 1991) acylations in organic media. The enzyme also catalyzes peptide synthesis, either by direct reversal of the hydrolytic process or by aminolysis of *N*-protected amino acid or peptide esters (Wong and Wang, 1991). Obtaining high enzyme activity is important for synthetic applications such as these. The *specific* activity of subtilisin E was greatly enhanced by the directed evolution experiment described above. However, the *total* activity of the *Bacillus subtilis* culture supernatant was reduced to <10% that of the starting wild type culture supernatant in the absence of DMF (Table I). This drop in total activity reflects a large decrease in the expression of the 'evolved' 10M enzyme in the *B.subtilis* host. Such changes in expression levels are not unexpected, since DNA base and amino acid substitutions may affect the transcription, translation or enzyme export and

© Oxford University Press

Downloaded from https://academic.oup.com/peds/article/9/1/77/1525399 by guest on 20 August 2022

maturation process. This result can be directly attributed to the use of a screening method that focused only on improving specific activity in aqueous DMF.

While mutations which slightly improved specific subtilisin activity in DMF could be found by screening only a few hundred colonies on average, fine-tuning a combination of expression levels and specific activity could be expected to require a search of larger variant libraries. Small libraries of subtilisin variants in *B.subtilis* can be obtained by direct transformation of the ligation products of the PCR random mutagenesis (Chen and Arnold, 1993). The low efficiency of *B.subtilis* transformation with ligated recombinant plasmids, however, limits the size of the mutant library that can be established in this host. Here we report further directed evolution of subtilisin, using a *B.subtilis–Escherichia coli* shuttle vector to facilitate establishment of the mutant library in *B.subtilis*. By slightly altering the screening method to reflect the total subtilisin activity as well as activity in aqueous DMF, significant improvements in both measures were obtained.

Materials and methods

Restriction enzymes were purchased from Boehringer Mannheim (Indianapolis, IN). Reagents for oligonucleotide synthesis were purchased from Pharmacia (Piscataway, NJ). Succinyl-Ala-Ala-Pro-Phe-*p*-nitroanilide (s-AAPF-pNa) was obtained from Sigma (St Louis, MO). Nitrocellulose membranes were purchased from Schleicher & Schuell (Keene, NH). Biodyne membranes were obtained from Pall (Glen Cove, NY).

Bacterial strains and plasmids

Bacillus subtilis strain DB428 and *Bacillus* cloning vector pKWZ containing the subtilisin E gene (Park *et al.*, 1989) were kindly provided by R.Doi, University of California, Davis. Plasmid pUC19 and *E.coli* strain ER 1648 were from New England Biolabs (Beverly, MA). Plasmid pUB110 was from Bacillus Genetics Stock Center, Department of Biochemistry, The Ohio State University. Phagemids SKII+ and SKII– were purchased from Stratagene (La Jolla, CA). Competent *E.coli* ER1648 was prepared according to the standard Ca^{2+} protocol (Sambrook *et al.*, 1989). *Bacillus subtilis* DB428 was made competent for transformation as described previously (Chen and Arnold, 1993). *Escherichia coli* ER1648 cells containing plasmid pBE2 were grown on LB agar plates or LB broth supplemented with 20 μg/ml kanamycin. *Bacillus subtilis* DB428 cells harboring plasmids pKWZ or pBE2 were grown on LB agar plates or LB broth supplemented with 50 μg/ml kanamycin.

Table I. Total activities of supernatants of *B.subtilis* expressing wild type and variant subtilisins E

Subtilisin E	0% DMF (units/ml)	20% DMF (units/ml)
WT	4.5	0.2
10M	0.27	0.21
12M	1.4	1.55
13M	2.7	3.4

Assay is for hydrolysis of s-AAPF-pNA in 0.1 mM Tris–HCl, 10 mM $CaCl_2$, pH 8.0 with and without 20% DMF (v/v).

Construction of B.subtilis–E.coli shuttle expression vector pBE1

Shuttle vector pBE1 was constructed by ligation of three DNA fragments. The first was the *Sna*BI–*Bam*HI fragment (nucleotides 2164–793) of plasmid pUB110 containing the *Bacillus* replication origin region and the kanamycin resistance gene (McKenzie *et al.*, 1986, 1987). The second and third fragments were generated from pUB110 and pUC19 by the PCR (Saiki *et al.*, 1988) using synthetic oligonucleotide primers. To obtain the second fragment, oligonucleotides MO5B (5′ CTCTTCCTTC GGATCCTATG TAAATCGCTC CTT 3′) and MO3P (5′ CCTCCTTTCC CAGCTGGAAC GAGACTTTGC AGT 3′) were used as the 5′ and 3′ primers, respectively, to amplify the region from nucleotides 1075 to 1651 of pUB110 containing the minus-strand replication origin. The 3′ primer was designed to eliminate the unique *Nde*I site of pUB110 to facilitate the subsequent cloning of the subtilisin E gene for random mutagenesis. The first 10 nucleotides at the 5′ ends of both primers were random sequences to facilitate restriction enzyme digestion. The underlined sequences were the restriction sites for *Bam*HI and *Pvu*II respectively. The third fragment containing the replication origin of *E.coli* plasmid pUC19 (nucleotides 630–1559) was generated using the 5′ and 3′ primers COE5H (5′ TCTTCAGCCTGTTGACCTGC ATTAATGAATCGG 3′) and COE3H (5′ TTCTCTTCCTGTTGACT**AAAAAGAAGCAGGTTTTCTTATACCTGCTTCTTTTTACAGCTG**ATCTAGGTGAAGATCCC 3′), respectively. The first 10 nucleotides of COE5H and COE3H are random sequences. The underlined sequences are restriction sites for *Hinc*II and *Pvu*II respectively. Primer COE3H contains the transcriptional terminator sequence of the subtilisin E gene (shown in bold face), which was designed to be positioned immediately downstream from the kanamycin gene in pBE1. The *Pvu*II site in primer COE3H was used to determine the orientation of the third fragment in pBE1 and to facilitate future plasmid manipulation. The subtilisin E gene was removed from plasmid pKWZ by cutting with *Eco*RI and *Bam*HI and subcloned into *Eco*RI–*Bam*HI-digested pBE1 to form pBE2 (Figure 1).

PCR-based random mutagenesis

Random mutagenesis was carried out using error-prone PCR (Leung *et al.*, 1989). Two synthetic oligonucleotides SUB5N (5′ GATCCGAGCGTTGCATATGTGGAAGAAGATCAT 3′) and SUB3B (5′ GGTTCTTGGATCCGATT-CAA CATGCGG-AG 3′) permitted amplification of the full-length subtilisin E sequence. The underlined sequences are *Nde*I and *Bam*HI restriction sites for SUB5N and SUB3B respectively, which allowed the PCR products to be ligated with vector pBE2 digested with the same enzymes. The ligation mixture was then transformed into *E.coli* ER1648 competent cells to generate a library of subtilisin E mutant genes. The mutant library was isolated from *E.coli* ER1648 host cells and transferred into *B.subtilis* DB428 competent cells for expression and screening.

Screening for enhanced subtilisin activity in aqueous DMF

Modified Schaefer's agar plates (Leighton and Doi, 1971) containing 1% casein were used for *B.subtilis* growth and subtilisin expression. Agar plates were first covered with two membranes, one organic solvent-resistant Biodyne nylon membrane and one nitrocellulose membrane, onto which *B.subtilis* transformants were plated. *Bacillus subtilis* cells containing an unmutated subtilisin E gene transformed with

Downloaded from https://academic.oup.com/peds/article/9/1/77/1525399 by guest on 20 August 2022

pBE2 were used as controls. After overnight incubation at 37°C, observable halos indicated that colonies were secreting active subtilisin. The top membrane carrying the *B.subtilis* colonies was removed and kept at 4°C. The bottom membrane (which contains enzyme at each spot corresponding to individual halos) was then transferred onto a new set of agar plates containing 1% casein and 35–45% DMF. These DMF plates were incubated at 37°C overnight. The halo sizes on both the aqueous and DMF plates were then evaluated and compared with those of unmutated control colonies. Colonies which outperformed the controls were selected and grown in 2 ml modified Schaeffer's medium at 37°C for 48 h in a rotary shaker. The supernatants of the liquid cultures were used for assays of total hydrolytic activity towards the peptide substrate s-AAPF-pNA in 0.1 mM Tris–HCl, pH 8.0 and 10 mM $CaCl_2$. DMF was added as indicated (volume %).

Sequencing of subtilisin E variant genes

Selected plasmids were digested with *Eco*RI and *Bam*HI and the subtilisin E gene subcloned into Bluescript SKII+ and SKII– phagemids with *Eco*RI and *Bam*HI. Single-stranded phagemids containing both orientations of a variant subtilisin E gene were prepared according to the supplier's instructions. Dideoxynucleotide DNA sequencing was carried out using the Sequenase II kit (USB).

Enzyme purification and kinetics

Subtilisin E purification and kinetic studies were carried out as described previously (Chen and Arnold, 1993). s-AAPF-pNa (0.24 mM) was used to assay hydrolytic activity in 0.1 M Tris–HCl, pH 8.0, 10 mM $CaCl_2$, 37°C, with DMF added as indicated. With the exception of the total activity assays, which were carried out using culture supernatant, all activity assays were carried out with purified enzymes. One unit of subtilisin E is the amount of enzyme required to catalyze the hydrolysis of 1 μmol substrate to products in 1 min.

Molecular modeling

Modeling of subtilisin E and its variants was carried out using the InsightII computer program based on the X-ray crystal structure of the subtilisin from *Bacillus mesentericus* (Dauter *et al.*, 1991) (Protein Data Bank reference: 1MEE). Subtilisin E from *B.subtilis* differs from subtilisin from *B.mesentericus* only at three out of 275 amino acid positions. The energy minimized s-AAPF-pNA–subtilisin complex is in good agreement with a similar model prepared for subtilisin BPN' and s-AAPF-pNa (Wells and Estell, 1988).

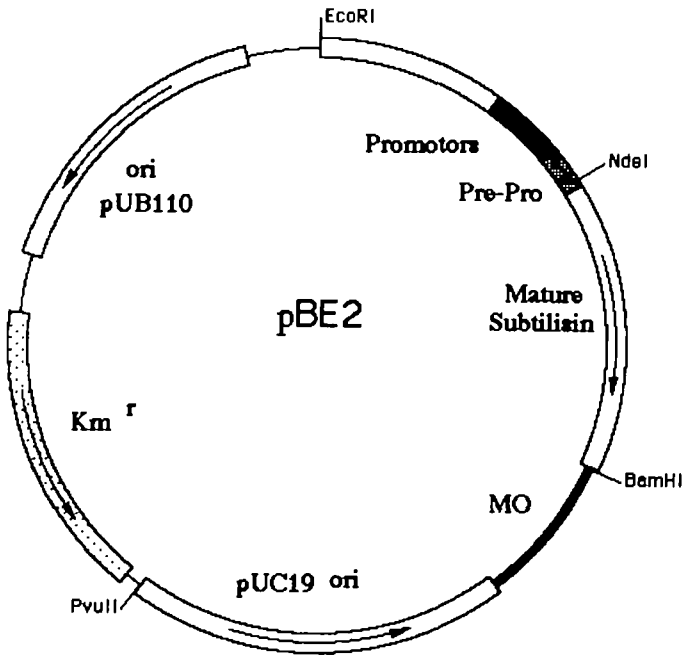

Fig. 1. The *B.subtilis–E.coli* shuttle vector pBE2 carrying a subtilisin E gene, constructed from pUB110 and pUC19 (see Materials and methods). The *Nde*I site of pUB110 downstream from the minus origin (MO) region was eliminated so that the *Nde*I site in the subtilisin E gene can be used for generating the PCR-based mutant library.

Results and discussion

The B.subtilis–E.coli shuttle expression vector pBE2 for directed evolution

Successful application of directed enzyme evolution requires screening large numbers of variants for the desired feature(s). Provided the screen is sufficiently sensitive, mutations conferring small enhancements can be identified and subsequently combined to achieve the desired result. The size of the mutant library that can be established in *E.coli* is generally not a limiting factor for screening, since even the most rapid screening methods can only handle ~10^6 or fewer clones. The direct cloning of mutant subtilisin E genes into *B.subtilis* competent cells, however, occurs with low efficiency (a few hundred transformants per microgram of DNA) relative to *E.coli* (10^6–10^7 transformants/μg DNA). Screening in *B.subtilis* is therefore quickly limited by the size of the mutant library. This is at least partially due to the fact that only multimeric linear or circular forms of plasmids can efficiently transform *B.subtilis* (Canosi *et al.*, 1978; Mottes *et al.*, 1979; de Vos *et al.*, 1981). The monomeric ligation products from the linearized plasmid and target gene cannot transform *B.subtilis* competent cells (de Vos *et al.*, 1981).

The *B.subtilis–E.coli* shuttle vector pBE1 was designed in order to facilitate the establishment of larger mutant libraries in *B.subtilis*. The shuttle vector was constructed by ligating

Table II. Kinetic parameters k_{cat}, K_m and k_{cat}/K_m for hydrolysis of S-AAPF-pNn by subtilisin E variants in 0.1 M Tris–HCl, 10 mM $CaCl_2$, pH 8.0, containing 0, 20 and 60% (v/v) DMF at 37°C

Variant	0% DMF			20% DMF			60% DMF	
	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m ($M^{-1}s^{-1}\times10^{-3}$)	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m ($M^{-1}s^{-1}\times10^{-3}$)	k_{cat}/K_m ($M^{-1}s^{-1}\times10^{-3}$)	
WT	0.56	21	38	17	12	1.4	0.014	(1)[a]
10M	0.1	27	270	0.7	73	99	2.2	(157)
13M	0.067	39	582	0.4	98	245	6.6	(471)

[a]Catalytic efficiency relative to wild type subtilisin E.

Downloaded from https://academic.oup.com/peds/article/9/1/77/1525399 by guest on 20 August 2022

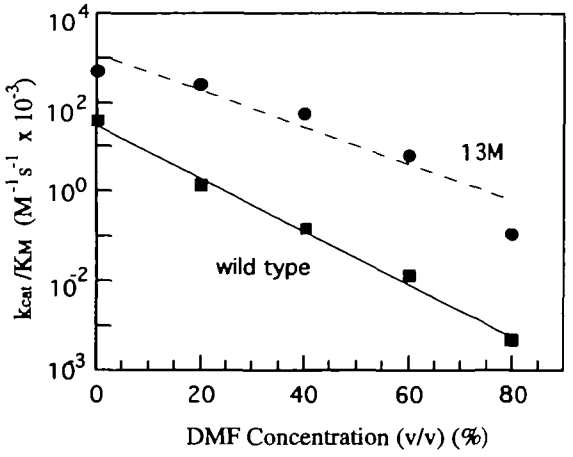

Fig. 2. Comparison of the catalytic efficiencies of wild type and 13M subtilisin E towards hydrolysis of s-AAPF-pNa in 0.1 M Tris–HCl, pH 8.0, 10 mM $CaCl_2$ with DMF (v/v), at 37°C.

three DNA fragments from plasmids pUB110 and pUC19 containing the replication origin region and the kanamycin resistance gene of pUB110, the portion of pUB110 covering the minus-strand replication origin and the replication origin region of pUC19 (see Materials and methods). The subtilisin E gene was subcloned from plasmid pKWZ to form pBE2 (Figure 1). The stabilities of plasmids pBE1 and pBE2 were found to be similar to that of pUB110 in *B.subtilis* (data not shown).

With the aid of shuttle vector pBE2, mutant libraries established first in *E.coli* could be easily amplified, isolated by a routine miniprep method and transferred into frozen *B.subtilis* competent cells. A significant fraction of plasmids isolated from *E.coli* ER 1648 is in the form of a supercoiled dimer of multimers (data not shown), which transform *B.subtilis* competent cells with high efficiency (de Vos *et al.*, 1981).

Screening for increased subtilisin activity in aqueous DMF

Sequential rounds of random mutagenesis and screening to identify and accumulate positive mutations resulted in a subtilisin E variant (10M) with significantly enhanced specific activity towards peptide hydrolysis in aqueous DMF (Chen and Arnold, 1993). The k_{cat}/K_m of the 10M variant in 60% DMF, for example, is enhanced 157-fold relative to the wild type enzyme for hydrolysis of s-AAPF-pNa (Table II). However, as shown in Table I, the total hydrolytic activity of the *B.subtilis* culture supernatant was reduced to less than 10% that of the starting wildtype culture, reflecting a very significant drop in the overall expression level. This is consistent with our observation that *B.subtilis* colonies expressing the 10M mutant generate much smaller halos on 1% casein plates containing no DMF than colonies expressing the wild type enzyme.

Starting with the gene for the 10M subtilisin E, two additional generations of PCR-based random mutagenesis and screening were conducted, using slightly modified screening criteria. After the first round of mutagenesis, one colony with a larger halo than the control colony (cells containing the parent 10M subtilisin E gene) on both casein plates containing DMF and no DMF was selected from several thousand colonies screened. That gene was isolated and subjected to a second round of random mutagenesis. A single colony producing larger halos on both aqueous and DMF plates was similarly identified from that *B.subtilis* library. The enzymes isolated from these first- and second-generation colonies are termed 12M and 13M, respectively.

Table III. Residual activities (reported as % of starting activity) of wild type, 10M and 13M subtilisins E after incubation in 70% DMF at 25°C

Subtilisin E	Incubation time		
	60 h	130 h	460 h
Wild type	91	88	48
10M	88	85	42
13M	92	89	50

Table IV. DNA and amino acid substitutions in 13M subtilisin E, obtained by site-directed mutagenesis and random mutagenesis and screening for enhanced specific activity in aqueous DMF

Base	Base substitution	Position in codon	Amino acid	Amino acid substitution
595	A→C	3	47	Silent
596	G→A	1		
597	C→G	2	48	Ala→Arg
632[a]	G→A	1	60	Asp→Asn
643	T→C	3	63	Silent
664	T→G	3	70	Silent
706	T→C	3	84	Silent
744[a]	A→G	2	97	Asp→Gly
762[a]	A→G	2	103	Gln→Arg
773	A→G	1	107	Ile→Val
838	A→G	3	128	Silent
846	G→A	2	131	Gly→Asp
871	T→C	3	139	Silent
921	A→G	2	156	Glu→Gly
946	A→G	3	164	Silent
996	A→G	2	181	Asn→Ser
998	A→G	1	182	Ser→Gly
1016	T→C	1	188	Ser→Pro
1033	T→C	3	193	Silent
1071	A→T	2	206	Gln→Leu
1107[a]	A→G	2	218	Asn→Ser
1126	T→C	3	224	Silent
1141	A→G	3	229	Silent
1153	A→G	3	233	Silent
1207	A→G	3	251	Silent
1216	A→G	3	254	Silent
1217	A→G	1	255	Thr→Ala

DNA and amino acid substitutions resulting from two generations of random mutagenesis and screening of 10M subtilisin E (Chen and Arnold, 1993) for enhanced total and specific activity in aqueous DMF are underlined.

[a]Substitutions incorporated by site-directed mutagenesis into DNA coding for 4M subtilisin E used as template for sequential random mutagenesis.

Catalytic activities of the 'evolved' enzymes

The total activities of supernatants of *B.subtilis* cultures expressing wild type, 10M and the further-evolved 12M and 13M subtilisin E genes towards hydrolysis of s-AAPF-pNa are summarized in Table I. The 12M variant from the first round of random mutagenesis showed significantly greater activity with and without DMF, relative to its parent 10M. The total activity is in fact slightly greater in the presence of 20% DMF than in purely aqueous buffer. The 13M variant from the second generation continues this trend: the total activity in the absence of DMF is 10-fold greater than the starting 10M variant, while activity in 20% DMF is improved by a factor of 16. In contrast to wild type subtilisin E, which

Downloaded from https://academic.oup.com/peds/article/9/1/77/1525399 by guest on 20 August 2022

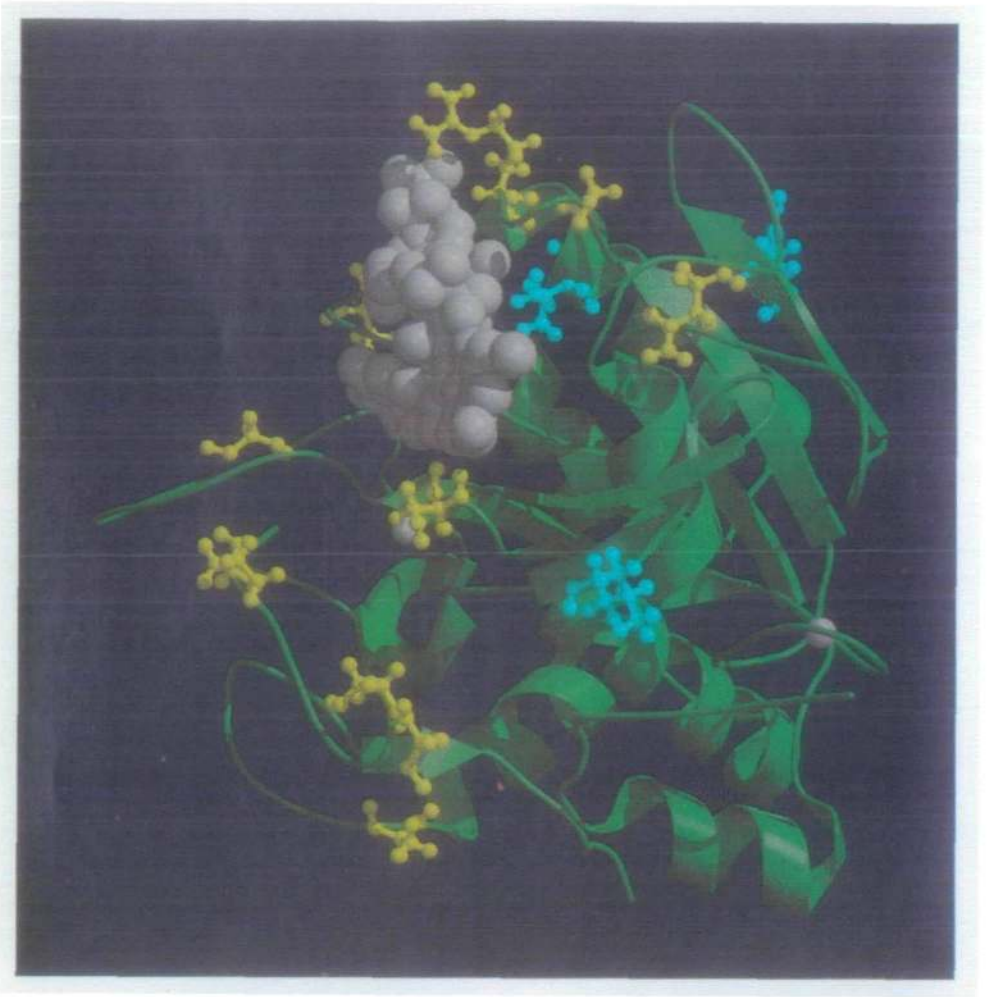

Fig. 3. Model of subtilisin E showing the 10 amino acid mutations in 10M (yellow) and the additional three mutations identified in the 13M variant (blue) evolved for increased total activity in DMF. Bound Ca^{2+} and peptide substrate s-AAPF-pNA are shown in gray.

retains only 4.4% of its activity in the presence of 20% DMF, 13M subtilisin E is actually 25% more active in 20% DMF than with no DMF at all.

The subtilisin variants were purified and further characterized. The kinetic parameters for the hydrolysis of s-AAPF-pNa by wild type subtilisin E and the selected variants are summarized in Table II. The specific catalytic efficiency k_{cat}/K_m of 13M subtilisin E is approximately 2.4-fold higher than that of 10M in 20% DMF and 175 times greater than that of the wild type enzyme. The increase in catalytic efficiency reflects an increase in k_{cat} as well as a decrease in the K_m (Table II).

As observed previously for the 10M variant, the enhancement in the catalytic efficiency of 13M subtilisin E extends to very high concentrations of DMF (Figure 2). The catalytic efficiency of 13M subtilisin is less sensitive to DMF than the wild type enzyme over the entire range of DMF concentration. In 60% DMF, the k_{cat}/K_m for 13M is 3-fold higher than 10M and 471-fold higher than the wild type subtilisin E.

The loss of activity that the wild type enzyme experiences in DMF has been substantially recovered by sequential generations of random mutagenesis and screening, to the point that the evolved enzyme is actually better at this particular catalytic task in low DMF concentrations (<40%) than the wild type enzyme in purely aqueous media. In 20% DMF, 13M subtilisin E is ~6.5 times more efficient than the wild type enzyme in aqueous solution (Figure 2). In 40% DMF, it is still 1.5 times more efficient than is wild type subtilisin E in the absence of DMF. This result is not as surprising as it might seem at first glance. It should be noted that in aqueous media the evolved enzymes are all 5- to 10-fold more efficient than the wild type enzyme. This improvement in the k_{cat}/K_m is largely due to a decrease in the K_m for the specific substrate s-AAPF-pNa used for screening (Chen and Arnold, 1993). Because the enzyme's evolution has been directed towards improving activity towards this particular substrate as well as catalytic activity in DMF, it is not unreasonable that the evolved enzyme is even more efficient than wild type.

Enzyme stability

The time courses of deactivation of the wild type, 10M and 13M subtilisins were compared at 25°C in 70% DMF (Table III). Under these conditions the evolved subtilisin will catalyze the polymerization of amino acid esters (Chen and Arnold, 1993). After 460 h incubation, mutant 13M retained ~50% activity, essentially the same as the wild type enzyme. As observed previously for the directed evolution of the 10M parent enzyme (Chen and Arnold, 1993), the mutations that enhance catalytic efficiency and total expressed activity have no significant net effect on the enzyme's stability, as measured

Downloaded from https://academic.oup.com/peds/article/9/1/77/1525399 by guest on 20 August 2022

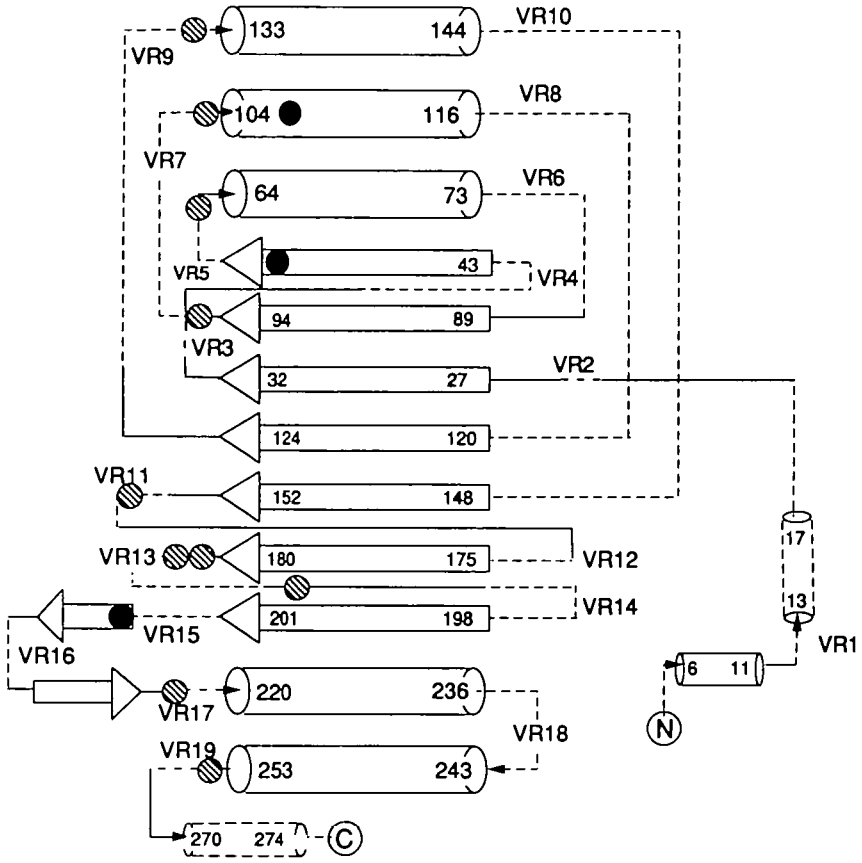

Fig. 4. Schematic representation of the secondary structure topology of subtilisin. α-Helices are represented as cylinders and β-sheets as arrows. Solid lines indicate conserved regions in subtilisin sequences. Dashed lines indicate variable regions (VRs) [redrawn from Siezen *et al.* (1991)]. Locations of amino acid substitutions in 10M subtilisin E are indicated by ○. The three additional substitutions in 13M are indicated by ●.

by residual activity after incubation in aqueous DMF. Since the enzyme is sufficiently stable to maintain its activity during the screening process, in the presence of DMF, obviously no benefit is derived from mutations that only enhance stability. The fact that mutations enhancing catalytic efficiency occur without corresponding changes in stability indicate that activity and stability in DMF are not coupled traits.

DNA and amino acid sequence changes

Site-directed mutagenesis followed by five generations of sequential random mutagenesis and screening resulted in the accumulation of a total of 27 DNA base substitutions coding for 13 amino acid substitutions in 13M subtilisin E (Table IV). Two generations of random mutagenesis and screening resulted in seven new base changes (underlined in Table IV) distributed across the 800 bp portion of the 10M subtilisin E gene sequence subjected to random mutagenesis. The PCR mutagenesis method used previously to obtain 10M subtilisin E involved introduction of dGTP and dATP in a 10:1 ratio and resulted in a preponderance of transitions (A→G or T→C). A more balanced distribution of transversions (three) and transitions (four) was obtained using the method of Leung *et al.* (1989). These seven base changes give rise to three amino acid substitutions, two of which (A48R and I107V) appeared during the first cycle of mutagenesis and screening (12M subtilisin E) and a third (Q206L) in the second cycle (13M). Substitutions I107V and Q206L are the results of single base substitutions in their corresponding codons. Substitution of Ala at position 48 by Arg, on the other hand, is the result of a double base change in the codon for amino acid 48.

As shown in Figure 3, the 10 amino acid substitutions in the parent 10M enzyme (shown in yellow) are clustered on the face of enzyme that harbors the active site and substrate binding pocket. The three new amino acid substitutions found in 13M subtilisin E (shown in blue) also appear more or less on the same face. All three new substitutions reside on the enzyme surface, as do all the previous 10 amino acid substitutions. Figure 3 shows that the 10 amino acid substitutions in 10M subtilisin E are located in loop structures connecting elements of secondary structure. In contrast, the three new amino acid substitutions in 13M are located in α-helix (I107V) or β-sheet (A48R and Q206L) secondary structures. I107V lies in the substrate binding pocket, where it interacts directly with the peptide substrate (shown in gray), while the remaining two mutations are located well outside this area. The substitution of Ile107 with Val was found in an earlier study to increase the stability slightly of subtilisin from *Bacillus amyloliquefaciens* towards autolysis under alkaline conditions (Cunningham and Wells, 1987).

Comparison of the amino acid sequences of more than 40 subtilisins from different sources shows a conserved core with insertions and deletions that are preferentially confined to surface loops (Siezen *et al.*, 1991). As indicated in Figure 4, the core α-helix and β-sheet secondary structures are relatively conserved, while the sequences of the peptide loops that connect these secondary structure elements are variable among

Downloaded from https://academic.oup.com/peds/article/9/1/77/1525399 by guest on 20 August 2022

the different subtilisins (VRs). Figure 4 clearly shows the positions of the 10 amino acid substitutions in 10M subtilisin E in these variable loops on the surface of the enzyme surrounding the active site. In contrast, the three new amino acid substitutions in 13M lie within secondary structures that are more conserved throughout the different subtilisins.

Although these secondary structures are more conserved in their sequences, variations do occur among subtilisins from different sources. Using the subtilisin sequences compared by Siezen *et al.* (1991), it can be seen that all three of the new amino acids in 13M are found in subtilisins from cyanobacteria (Arg48) and Gram-positive (Val107 and Leu206) or Gram-negative bacteria (Leu206). One subtilisin (from *Staphylococcus epidermis*) contains two of the three new amino acids in 13M (Val107 and Leu206); none of the sequences compared, however, have all three. Thus the individual 'solutions' found by this directed evolution approach have indeed been explored by nature, albeit for other purposes. As noted previously (Chen and Arnold, 1993), at least seven of the 10 amino acid substitutions in 10M subtilisin E are found in subtilisins from other sources. The new 13M subtilisin variant can therefore be thought of as largely a novel 'mix and match' of other, naturally occurring enzymes.

It is unknown how the mutations in the subtilisin gene reduce the expression of active enzyme in the 10M variant or why expression has been recovered in the 13M variant. Mutations at the DNA level can affect the transcription and translation processes, while mutations at the protein level can affect the folding and processing of the enzyme. For example, the subtilisin maturation process requires a pro-sequence which acts as an intramolecular chaperon for the correct folding of the enzyme (Ikemura *et al.*, 1987). Changes in local secondary or tertiary structures resulting from amino acid substitutions could interfere with the maturation process and reduce enzyme yield. The enzyme yield was (purposefully) not maintained during the earlier directed evolution for enhanced specific activity in DMF. The amino acid 'solutions' to this problem found in 10M subtilisin E all share the following features: they are located on the surface of the enzyme around the active site (although not necessarily adjacent) and they are all located in sequence-variable loops. When the screening criteria were modified in the current experiment to include an enhanced enzyme yield, the nature of the 'solutions' changed. That the three new amino acid substitutions in conserved secondary structures in 13M contrast with the positions of the 10 previous mutations in variable loops (Figure 4) hints at a mechanism for recovering total activity that occurs at the protein level.

A basic rule of directed evolution (and any random mutagenesis experiment) is, 'you get what you screen for'. The screen (or selection) should therefore reflect the desired result as closely as possible. In particular, the standards for selecting a positive should include all features of interest, in order to minimize unwanted side-effects such as a lower expression level or changes in other features markedly affected by mutations at the DNA or amino acid levels. These realities often work against the use of a selection scheme that offers the ability to search larger fractions of sequence space for something no-one wants.

Although screening limits the search for beneficial mutations to a very small fraction of sequence space, a workable strategy for directed evolution is to accumulate mutations conferring small enhancements in order to achieve the desired property (Arnold, in press). Sequential generations of error-prone PCR mutagenesis and screening is one effective method by which this can be accomplished. The accumulation of amino acid substitutions in an enzyme can result in significant improvements in performance, particularly for features not optimized under selective pressure.

Acknowledgements

The authors are grateful to Jeffrey Moore and Huimin Zhao for their assistance with the modeling and Figure 3. This work was supported by the US Department of Energy's program in Biological and Chemical Technologies Research within the Office of Industrial Technologies, Energy Efficiency and Renewables.

References

Arnold,F.H. (1993) *FASEB J.*, **7**, 744–749.
Arnold,F.H. (1996) *Chem. Eng. Sci.*, in press.
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.
Canosi,U., Morelli,G. and Trautner,T.A. (1978) *Mol. Gen. Genet.*, **166**, 259–267.
Chen,K.Q. and Arnold,F.H. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 5618–5622.
Cunningham,B.C. and Wells,J.A. (1987) *Protein Engng*, **1**, 319–325.
Dauter,Z., Betzel,C., Genov,N., Pipon,N. and Wilson,K.S. (1991) *Acta Crystallogr.*, **B47**, 707–730.
de Vos,W.M., Venema,G., Canosi,U. and Trautner,T.A. (1981) *Mol. Gen. Genet.*, **181**, 424–433.
Graham,L.D., Haggett,K.D., Jennings,P.A., Le Brocque,D.S. and Whittaker,R.G. (1993) *Biochemistry*, **32**, 6250–6258.
Ikemura,H., Takagi,H. and Inouye,M. (1987) *J. Biol. Chem.*, **262**, 7859–7864.
Joyet,P., Declerck,N. and Gailardin,C. (1992) *Bio/Technology*, **10**, 1579–1583.
Leighton,T.J. and Doi,R.H. (1971) *J. Biol. Chem.*, **246**, 3189–3195.
Leung,D.W., Chen,E. and Goeddel,D.V. (1989) *Technique*, **1**, 11–15.
Liao,H., McKenzie,T. and Hageman,R. (1986) *Proc. Natl Acad. Sci. USA*, **83**, 576–580.
McKenzie,T., Hoshiro,T., Tanaka,T. and Sueka,N. (1986) *Plasmid*, **15**, 93–103.
McKenzie,T., Hoshiro,T., Tanaka,T. and Sueoka,N. (1987) *Plasmid*, **17**, 83–85.
Margolin,A.L., Fitzpatrick,P.A., Dubin,P.L. and Klibanov,A.M. (1991) *J. Am. Chem. Soc.*, **113**, 4693–4694.
Mottes,M., Grandi,G., Sgaramella,V., Canosi,U., Morelli,G. and Trautner,T.A. (1979) *Mol. Gen. Genet.*, **174**, 281–286.
Oliphant,A.R. and Struhl,K. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 9094–9098.
Park,S.-S., Wong,S.-L., Wang,L.-F. and Doi,R.H. (1989) *J. Bacteriol.*, **6**, 521–527.
Riva,S. and Klibanov,A.M. (1988) *J. Am. Chem. Soc.*, **10**, 3291–3295.
Riva,S., Chopineau,J., Kieboom,A.P.G. and Klibanov,A.M. (1988) *J. Am. Chem. Soc.*, **110**, 584–589.
Saiki,R.K., Gelfand,D.H., Stoffel,S., Scharf,S.J., Higuchi,R., Horn,G.T., Mullis,K.B. and Erlich,H.A. (1988) *Science*, **239**, 487–491.
Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Siezen,R.J., De Vos,W.M., Leunissen,J.A.M. and Dijkstra,B.W. (1991) *Protein Engng.*, **4**, 719–737.
Stemmer,W.P.C. (1994) *Nature*, **370**, 389–391.
Strausberg,S.L., Alexander,P.A., Gallagher,D.T., Gilliland,G.L., Barnett,B.L. and Bryan,P.N. (1995) *Bio/Technology*, **13**, 669–673.
Wells,J.A. and Estell,D.A. (1988) *Trends Biochem. Sci.*, **13**, 291–297.
Wong,C.-H. and Wang,K.-T. (1991) *Experientia*, **47**, 1123–1129.

Received September 5, 1995; accepted on October 14, 1995

Downloaded from https://academic.oup.com/peds/article/9/1/77/1525399 by guest on 20 August 2022