J Mol Evol (1999) 49:716–720

JOURNAL OF MOLECULAR EVOLUTION

Exploring Nonnatural Evolutionary Pathways by Saturation Mutagenesis: Rapid Improvement of Protein Function

Kentaro Miyazaki, Frances H. Arnold

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

Received: 15 March 1999 / Accepted: 28 June 1999

Abstract. Random point mutagenesis does not access a large fraction of protein sequence space corresponding to primarily nonconservative amino acid substitutions. The cost of this limitation during directed evolution is unknown. Random point mutagenesis over the entire gene encoding the psychrophilic protease subtilisin S41 identified a pair of residues (Lys211 and Arg212) where mutations provided significant increases in thermostability. These were subjected to saturation mutagenesis to test whether the amino acids not easily accessible by point mutagenesis provide even better "solutions" to the thermostabilization challenge. A significant fraction of these variants surpassed the stability of the variants with point mutations. DNA sequencing revealed highly hydrophobic residues in the four most stable variants (Pro/Ala, Pro/Val, Leu/Val, and Trp/Ser). These nonconservative replacements, accessible only by multiple (two to three) base substitutions in a single codon, would be extremely rare in a point mutation library. Such replacements are also extremely rare in natural evolution. Saturation mutagenesis may be used advantageously during directed evolution to explore nonnatural evolution pathways and enable rapid improvement in protein traits.

Key words: directed evolution — fitness — random mutagenesis — saturation mutagenesis — subtilisin — thermostability

Correspondence to: F.H. Arnold; *e-mail:* frances@cheme.caltech.edu

Introduction

Directed evolution has emerged in the last several years as a powerful tool for altering and improving protein functions. The screening (or selection) of genetic diversity created by random mutagenesis and/or recombination can rapidly generate desired functional changes (Chen and Arnold 1993; Stemmer 1994; Arnold 1998). Widely used in directed evolution experiments are error-prone PCR (Leung et al. 1989; Chen and Arnold 1993) and other mutagenesis methods that generate single base mutations. Because mutation rates must be low for whole-gene evolution in order to observe rare beneficial mutations, a large fraction of protein sequence space is inaccessible in any one mutation operation. With single base mutations, only 5.7 amino acid substitutions on average are accessible from any given amino acid residue. A further consequence of the structure of the genetic code is that single base changes usually generate conservative amino acid substitutions, replacing one amino acid with others having similar physicochemical properties.

Little is known about the cost (or benefit) of limiting the search space in this way during directed evolution (Del Rio et al. 1994; Sirotkin 1986). There may be little cost to not exploring nonconservative substitutions. It is believed that natural selection works against drastic changes in amino acid properties to maintain protein functions (Schulz and Schirmer 1979). Mutation data matrices that summarize amino acid substitution tendencies in homologous proteins also seem to support this rule (Dayhoff et al. 1978; George et al. 1990). Common guidelines for site-directed mutagenesis experiments designed to probe or modify protein functions recommend

substitution with similar amino acids (Bordo and Argos 1991). However, these observations are all based on the study of natural proteins and therefore are biased by the existence of many more functionally neutral mutations than adaptive ones.

Directed evolution experiments, in contrast, accumulate mutations that are for the most part adaptive (Giver et al. 1998). Therefore, the mutation patterns will not necessarily resemble those found in natural sequences. We wished to explore how expanding the search space affects directed protein evolution by introducing the potential for more drastic amino acid changes. For this study, we targeted the thermostability of a psychrophilic enzyme, subtilisin S41 (Devail et al. 1994). This enzyme is significantly less stable than its mesophilic and thermophilic homologs; its directed evolution offers a convenient test of the adaptive process. Due to the absence of a convenient method for codon-based random mutagenesis that can be applied to the whole gene, we have used saturation mutagenesis to access all possible amino acid substitutions at two residues believed to offer opportunity for improvement in fitness.

Materials and Methods

Random Mutagenesis

Subtilisin S41 was expressed using plasmid pCT1, which carries the prosubtilisin sequence of S41 fused to presequence of subtilisin BPN′ (Miyazaki et al. in preparation). The pCT1 was created from pPG2267 (provided by Dr. Rowan Grayling, Procter & Gamble) by adding a *Bsu*36I site in front of the mature S41 sequence. Error-prone PCR was carried out using primers Bsu36I-1 (5′-CAAACAAACCTGAGGCTCTTTACAACGC-3′) and DHin (5′-ATTACGCCAAGCTTGGATCCTTA-3′), which allow random mutagenesis of the entire S41 mature sequence. A reaction contained 10 m*M* Tris-HCl (pH 8.3 at 25°C), 50 m*M* KCl, 7 m*M* $MgCl_2$, 0.15 m*M* $MnCl_2$, 0.01% (w/v) gelatin, 0.2 m*M* dATP, 0.2 m*M* dGTP, 1 m*M* dCTP, 1 m*M* dTTP, 50 pmol of each primer, 10 ng of template plasmid, and 5 U *Taq* polymerase (Promega) in a total volume of 100 μl. PCR was carried out on an MJ Research thermal cycler (PTC-200) at 94°C for 30 s, 55°C for 30 s, and 72°C for 30 s, and a total of 30 cycles were performed. These conditions should generate an error frequency of approximately two to three substitutions per 1,000 bases, or approximately one amino acid substitution per gene copy (Leung et al. 1989; Zhao et al. 1999).

Saturation Mutagenesis

Saturation mutagenesis (Hayashi et al. 1994) at amino acid positions 211 and 212 was carried out using a set of degenerate synthetic oligonucleotides KR+ (5′-CGT GGC CAT (AGCT)(AGCT)(GT) (AGCT)(AGCT)(GT) ACT GCT GGA GAT TAC GTT ATC-3′) and KR− (5′-TCC AGC AGT (AC)(AGCT)(AGCT) (AC)(AGCT)(AGCT) ATG GCC ACG ACA TGA AAA GTC-3′). Briefly, the upstream gene fragment (amino acids 1–219) was amplified with *Bsu*36I-1 and KR−, and the downstream fragment (amino acids 203–309) was amplified with KR+ and DHin. The amplified fragments were purified and mixed in equal amounts. The full-length S41 gene (amino acids 1–309) was then amplified with Bsu36I-1 and DHin.

Library Production and Screening

Products from the mutagenic PCR reaction were purified using a Qiagen DNA purification kit and cloned back into expression vector pCT1 using *Bsu*36I and *Bam*HI restriction sites. The insert fragment was cut with *Bsu*36I (New England Biolabs) and *Bam*HI (Boerhinger Mannheim) simultaneously in 50 m*M* Tris-HCl (pH 7.5), 10 m*M* $MgCl_2$, 0.1 *M* NaCl, and 1 m*M* dithiothreitol at 37°C for 2 h. The resulting DNA was then run on a 0.8% agarose gel and the ~1 kbp band was excised from the gel and extracted using a QIAquick kit (Qiagen). Purified DNA was eluted in water. Ligation reactions were performed using T4 DNA ligase (Stratagene). Vector DNA (the entire pCT1 excluding the mature region of S41 between *Bsu*36I and *Bam*HI), insert DNA (the S41 gene between *Bsu*36I and *Bam*HI), 10× ligation buffer, 1 m*M* ATP, water, and enzyme were combined and incubated at room temperature overnight.

Resulting plasmids were introduced into *Escherichia coli* HB101 competent cells (Sambrook et al. 1989), and cells were plated on LB agar plates containing 100 μg/ml ampicillin. Plasmids were purified from colonies using a Qiagen mini-prep kit. The plasmid library was used to transform *Bacillus subtilis* DB428 competent cells (Harwood and Cutting 1990), which were plated on LB agar plates containing 25 μg/ml kanamycin. Colonies appearing after transformation were picked with a sterile toothpick and resuspended in separate wells of a 96-deep-well plate (Bel-Art Products) containing 1 ml of 2xYT (1.6% tryptone, 1.0% yeast extract, 0.5% NaCl), 10 m*M* $CaCl_2$, and 25 μg/ml kanamycin. The plates were shaken at 300 rpm at 37°C for 48 h. After centrifugation at 4,000 *g* for 5 min at 20°C, a 10-μl aliquot from each well was pipetted into a second 96-well plate (Rainin), to which was added 100 μl of a substrate solution containing 0.1 *M* Tris-HCl (pH 8.5), 10 m*M* $CaCl_2$, 0.2 m*M* succinyl-LAla-LAla-LPro-LPhe-*p*-nitroanilide (s-AAPF-pNA) (Sigma). The resulting reaction was monitored at 25°C on a Molecular Devices plate reader. Relative reaction rates were determined from the rate of change in absorbance at 405 nm.

Protein Purification

Wild-type S41 and variants were purified from *B. subtilis* culture supernatants by ammonium sulfate preciptation (70% saturation) and chromatography. The precipitate was recovered by centrifugation at 15,000 *g* for 30 min at 4°C. The pellet was resuspended in 50 m*M* Tris-HCl (pH 8.5), 10 m*M* $CaCl_2$, and 1.5 *M* ammonium sulfate, and the protein was then applied to a RESOURCE® PHE (Pharmacia) column (1 ml, 6.4 mm ID × 3 cm) pre-equilibrated with buffer. The column was rinsed with 5 ml buffer, and the proteins were eluted with a linear gradient of ammonium sulfate to 0 *M*. Active fractions were pooled and concentrated to ~1 ml using an Amicon YM-10 membrane equipped in an ice-cold ultrafiltration unit. The sample was then applied to a HiPrep® Sephacryl S-100 16/60 (Pharmacia) gel filtration column (120 ml, 26 mm ID × 60 cm) pre-equilibrated in 50 m*M* Tris-HCl (pH 8.0), 10 m*M* $CaCl_2$, and 1.5 *M* NaCl. Active fractions were combined and dialyzed against 50 m*M* Tris-HCl (pH 8.0) and 10 m*M* $CaCl_2$.

Enzyme Activity and Stability

Proteolytic activity was determined in 0.1 *M* Tris-HCl (pH 8.5), 10 m*M* $CaCl_2$, and 0.2 m*M* s-AAPF-pNA at 30°C in a thermostatted Shimadzu BioSpec-1601 spectrophotometer. The reaction was followed by monitoring the formation of *p*-nitroaniline at 410 nm (DelMar et al. 1979). Half-lives of thermal inactivation at 60°C were determined in 50 m*M* Tris-HCl (pH 8.0) and 10 m*M* $CaCl_2$ as described (Zhao and Arnold 1999).

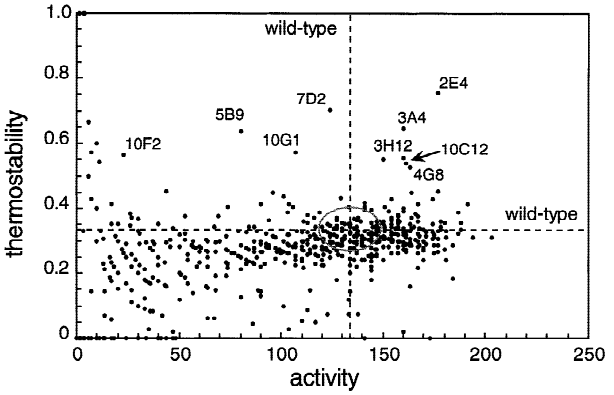

Fig. 1. Activities and thermostabilities of S41 variants prepared by PCR mutagenesis. Stability is measured by the ratio of residual activity following incubation at 60°C for 15 min to initial activity. Activity was measured in 0.1 *M* Tris-HCl (pH 8.5), 10 m*M* $CaCl_2$, and 0.2 m*M* s-AAPF-pNA at 25°C. *Ellipse* indicates standard deviations in activity and thermostability values for 96 wild-type clones. The inherent variability of the screen is small compared to the variability due to mutation.

Table 1. Mutations found in thermostable subtilisin S41 variants

Enzymes	Mutations
PCR mutagenesis	
1-2E4	K211E [A631G][a]
1-3A4	S145I [G434T]
1-3H12	S295T [T883A]
1-4G8	K221E [A661G], [t291c]
1-5B9	R212C [C634T]
1-7D2	F60L [T178C], K211E [A631G]
1-10C12	S175T [T523A], [a60g]
1-10F2	S138T [A412T], K211N [A633C]
1-10G1	N15D [A43G], [a240g]
Saturation mutagenesis	
wild type	K211 [AAA]/R212 [CGT][b]
1-2E1	W211 [TGG]/S212 [TCT]
1-8F1	P211 [CCT]/V212 [GTT]
1-9H9	L211 [CTC]/V212 [GTG]
1-14A7	P211 [CCG]/A212 [GCT]

[a] Base substitutions are shown in brackets. Nonsynonymous mutations are indicated in capital letters; synonymous mutations are given by lowercase letters.
[b] Codon sequences.

Results and Discussion

Subtilisin S41 was subjected to random mutagenesis by error-prone PCR under conditions expected to yield two to three base substitutions per gene. A total of 864 clones were picked and screened for activity before and after incubation at 60°C for 15 min. Ninety-six wild-type clones were also picked and subjected to the same screening procedure. The results are shown in Fig. 1. The inherent variability of the screen is reflected in the distribution of stability and activity values for the wild-type clones (indicated by the standard deviation ellipse in Fig. 1). Sources of error, which include well-to-well variations in growth conditions and pipetting errors, were minimized to obtain this low level of variation. Nine clones showing thermostability well above this wild-type region were picked for further characterization. Of these, eight show activity comparable to wild type, and one (1-10F2) has much lower activity.

All the genes encoding thermostable variants contained mutations that alter the S41 amino acid sequence (Table 1). Three of the variants (1-2E4, 1-7D2, and 1-10F2) have amino acid substitutions at position 211, and a fourth (1-5B9) contained a mutation that alters residue 212. The reduced activity of 1-10F2 may be caused by a second mutation that alters S138, a position that is completely conserved in all subtilisin sequences so far reported (Siezen and Leunissen 1997). The region containing residues 211 and 212 appears to be a weak point in this psychrophilic enzyme.

We then used saturation mutagenesis to explore whether there might be more thermostable sequences than those discovered in the point mutation library. A total of 1,536 clones from the library produced by saturation mutagenesis at the codons for residues 211/212 were picked and screened for proteolytic activity. As shown in Fig. 2A, a large fraction of the library loses activity—the region is quite sensitive to amino acid replacement. One hundred and five clones showing reasonable activity were assayed for thermostability by measuring their residual activities following incubation at 60°C for 20 min. The stabilities were widely distributed, with a large number of variants showing higher stability than wild type (Fig. 2B). There is no correlation between enzyme activity and thermostability in this library. The four most thermostable variants indicated in Fig. 2B were purified to homogeneity, and their thermostabilities were compared to wild type. All four variants were more thermostable than those isolated from the point mutagenesis library. Variant 1-14A7 showed the highest thermostability, with a half-life at 60°C of 84 min, versus 27–30 min for enzymes from the point mutagenesis library and 8 min for wild type.

DNA sequencing of these four most stable variants revealed the following thermostable sequences at 211/212: Trp/Ser, Pro/Val, Leu/Val, and Pro/Ala (Table 1). In all cases, the region was occupied by highly hydrophobic residues very unlike the wild-type sequence of Lys/Arg. Multiple mutations were required for these replacements. These residues are located in an extended mobile loop that has been proposed to be responsible for the high specific activity of this psychrophilic enzyme while compromising its stability (Devail et al. 1994). The sequences in this region are not conserved in the closely related subtilisins that also also have this extended loop: psychrophilic *Bacillus* TA39 (Ser/Arg) (Narinx et al. 1992) and mesophilic *B. sphaericus* (Ile/Ser) (Wati et al. 1997).

Mutation data matrices, which summarize patterns and frequencies of amino acid substitutions in natural

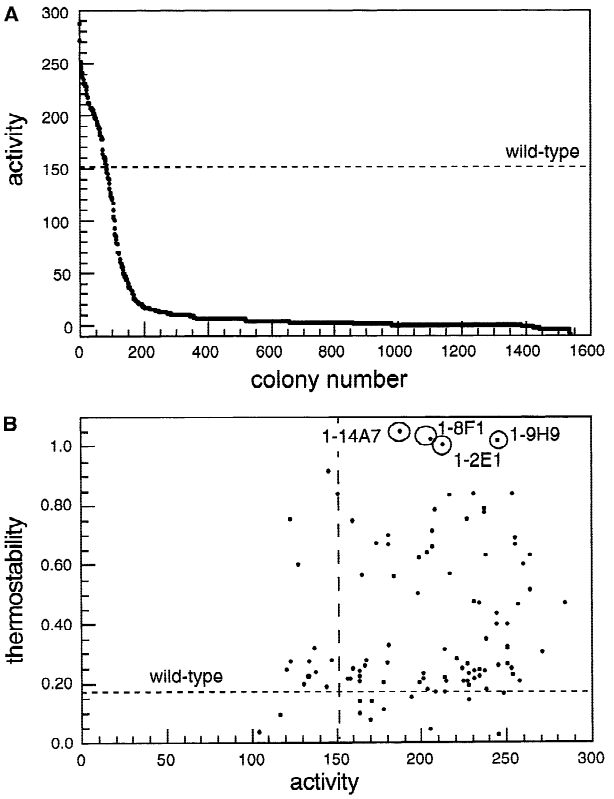

Fig. 2. **A** Relative activities of S41 variants from library prepared by saturation mutagenesis at residues 211/212, plotted in descending order. **B** Activity versus thermostability of active variants. Four most thermostable variants are indicated.

proteins (Dayhoff et al. 1978; George et al. 1990), indicate that (1) proteins are highly mutable; (2) protein surfaces are more tolerant to substitutions than interiors; and (3) some residues (e.g., Asn, Ser, Asp, and Glu) are highly mutable, whereas others (Trp and Cys) are rarely changed. However, most mutations are functionally neutral, or very nearly so (Kimura 1983). In directed evolution, very strong selective pressures are applied in a process that generates primarily adaptive mutations. Recent efforts to direct the evolution of mesophilic enzymes to create their thermophilic counterparts, for example, showed that large differences in thermostability (14–18°C) could be obtained in relatively few generations, with sequences that are 95–97% identical (Giver et al. 1998; Zhao and Arnold 1999). In contrast, natural mesophilic and thermophilic homologs are often separated by hundreds of amino acids.

Molecular evolution tends not to access large portions of protein sequence space corresponding to the mainly nonconservative amino acid substitutions that require multiple base substitutions within a single codon. It is possible that most of the sequence space that is not accessed (mainly nonconservative amino acid substitutions) would not contribute to the adaptive process and in fact would reduce the efficiency of directed evolution by diluting the libraries with large numbers of less fit sequences. At least in this example, however, the nonconservative substitutions offered greater potential for fitness gains. Nonconservative substitutions may in fact be a rich genetic resource for adaptation to new functional challenges. In the absence of a good method for codon-based mutagenesis, saturation mutagenesis at sites identified during whole-gene mutagenesis to have potential for improvement may prove a valuable addition to the directed evolution toolbox.

Acknowledgments. This research is funded by Procter & Gamble. We wish to thank Rowan Grayling and Donn Rubingh for materials and helpful discussions.

References

Arnold FH (1998) Design by directed evolution. Acc Chem Res 31: 125–131

Bordo D, Argos P (1991) Suggestions for “safe” residue substitutions in site-directed mutagenesis. J Mol Biol 217:721–729

Chen K, Arnold FH (1993) Tuning the activity of an enzyme for unusual environments: sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide. Proc Natl Acad Sci USA 90:5618–5622

Dayhoff MO, Schwartz RM, Orcutt BC (1978) A model of evolutionary change in proteins. In: Dayhoff MO (ed) Atlas of protein sequence and structure. National Biomedical Research Foundation, Washington, DC, pp 345–352

DelMar EG, Largman C, Brodrick JW, Geokas MC (1979) A sensitive new substrate for chymotrypsin. Anal Biochem 99:316–320

Del Rio G, Osuna J, Soberon X (1994) Combinatorial libraries of proteins: analysis of efficiency of mutagenesis techniques. Biotechniques 17:1132–1139

Devail S, Feller G, Narinx E, Gerday C (1994) Cold adaptation of proteins—purification, characterization, and sequence of the heat-labile subtilisin from the Antarctic psychrophile *Bacillus* TA41. J Biol Chem 269:17448–17453

George DG, Baker WC, Hunt LT (1990) Mutation data matrix and its uses. Methods Enzymol 183:333–351

Giver L, Gershenson A, Freskgard PO, Arnold FH (1998) Directed evolution of a thermostable esterase. Proc Natl Acad Sci USA 95:12809–12813

Harwood CR, Cutting SM (1990) Molecular biological methods for *Bacillus.* John Wiley and Sons, England

Hayashi N, Welschof M, Zewe M, Braunagel M, Dubel S, Breitling F, Little M (1994) Simultaneous mutagenesis of antibody CDR regions by overlap extension and PCR. Biotechniques 17:310–315

Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, Cambridge

Leung DW, Chen E, Goeddel DV (1989) A method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction. Technique 1:11–15

Miyazaki K, Wintrode PL, Grayling RA, Rubingh DN, Arnold FH, (in press) Directed evolution study of temperature adaptation in a psychrophilic enzyme

Narinx E, Devail S, Feller G, Gerday C (1992) Nucleotide and derivative amino acid sequence of the subtilisin from the Antarctic psychrotroph *Bacillus* TA39. Biochim Biophys Acta 1131:111–113

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2d ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Schulz GE, Shirmer RH (1979) Principles of protein structure. Springer-Verlag, New York, NY

Siezen RJ, Leunissen JAM (1997) Subtilases: the superfamily of subtilisin-like serine protease. Protein Sci 6:501–523

Sirotkin K (1986) Advantages to mutagenesis techniques generating populations containing the complete spectrum of single codon changes. J Theor Biol 123:261–279

Stemmer WP (1994) Rapid evolution of a protein in vitro by DNA shuffling. Nature 370:389–391

Wati MR, Thanabalu T, Porter AG (1997) Gene from tropical *Bacillus sphaericus* encoding a protease closely related to subtilisins from Antarctic bacilli. Biochim Biophys Acta 1352:56–62

Zhao H, Arnold FH (1999) Directed evolution converts subtilisin E into a functional equivalent of thermitase. Protein Eng 12:47–53

Zhao H, Moore JC, Volkov AA, Arnold FH (1999) Methods for optimizing industrial enzymes by directed evolution. In: Demain AL, Davies JE (eds) Manual of industrial microbiology and biotechnology. American Society for Microbiology, Washington, DC, pp 597–604