microbial biotechnology

Microbial Biotechnology (2011) **4**(6), 767–776 doi:10.1111/j.1751-7915.2011.00279.x

Enzymes for the laundry industries: tapping the vast metagenomic pool of alkaline proteases

F. Niehaus,[1]* E. Gabor,[1] S. Wieland,[2] P. Siegert,[2] K. H. Maurer[2] and J. Eck[1]

[1]*BRAIN AG (Biotechnology Research and Information Network), Darmstädter Straße 34-36, D-64673 Zwingenberg, Germany.*

[2]*Henkel AG & Co.KGaA, Henkelstrasse 67, D-40191 Duesseldorf, Germany.*

Summary

In the wide field of laundry and cleaning applications, there is an unbroken need for novel detergent proteases excelling in high stability and activity and a suitable substrate range. We demonstrated the large amount of highly diverse subtilase sequences present in metagenomic DNA by recovering 57 non-redundant subtilase sequence tags with degenerate primers. Furthermore, an activity- as well as a sequence homology-based screening of metagenomic DNA libraries was carried out, using alkaline soil and habitat enrichments as a source of DNA. In this way, 18 diverse full-length protease genes were recovered, sharing only 37–85% of their amino acid residues with already known protease genes. Active clones were biochemically characterized and subjected to a laundry application assay, leading to the identification of three promising detergent proteases. According to sequence similarity, two proteases (HP53 and HP70) can be classified as subtilases, while the third enzyme (HP23) belongs to chymotrypsin-like S1 serine proteases, a class of enzymes that has not yet been described for the use in laundry and cleaning applications.

Introduction

Enzymes are increasingly used as active ingredients in consumer products, detergent enzymes being a very prominent example. Particularly proteases are well-established constituents of modern washing and cleaning products, allowing the removal of protein-containing soiling. Due to their stability at high pH and temperature and their tolerance towards elevated concentrations of denaturing agents such as detergents or oxidants, subtilases are mostly used in this field of application. Subtilases are non-specific serine endopeptidases (superfamily S8 according to the Merops system of classification, http://merops.sanger.ac.uk) and can be divided into six families by sequence alignment (Siezen and Leunissen, 1997). Commercially relevant enzymes group into true and high-alkaline subtilisins (family A), Thermitases (family B) and Proteinase K-type enzymes (family C). Family A comprises the widely used detergent proteases of *Bacillus licheniformis* [subtilisin Carlsberg (Smith *et al.*, 1968; Jacobs *et al.*, 1985)], *B. amyloliquefaciens* [subtilisin BPN' (Wells *et al.*, 1983; Vasantha *et al.*, 1984)] and *B. lentus* [subtilisin BL (Goddette *et al.*, 1992)]. During the past decades, many efforts have been made to identify new proteases and optimize existing ones to meet the specific requirements for very diverse laundry and cleaning processes (Herbots, 2007).

Industrially relevant proteases so far have been obtained by screening culture collections or by using classical microbial enrichment techniques. However, this approach is intrinsically limited to microorganisms that are able to grow under laboratory conditions, thereby only capturing a small fraction of the microbial diversity present in nature. It has been estimated that less than 1% of microorganisms in the environment can be cultivated by using standard laboratory techniques (Amann *et al.*, 1995). Due to this general limitation of traditional culture-based enzyme recovery, the screening of metagenomic DNA libraries has become an attractive alternative. It does not require the cultivation of microorganisms, but solely relies on the genetic information stored in the collective genomes of all microorganisms present in an environmental sample, the so-called metagenome (Handelsman *et al.*, 1998). During the past years, a variety of new enzymes has been recovered by this method (for a recent compilation see Eck *et al.*, 2009), usually being only distantly related to sequence database entries.

In this study, we evaluated metagenome screening for the isolation of novel proteases for laundry and cleaning applications. Using degenerate primers that target family

Received 16 December, 2010; accepted 20 May, 2011. *For correspondence. E-mail fn@brain-biotech.de; Tel. (+49) 6251 9331 0; Fax (+49) 6251 9331 11.

A (and a subgroup A-alk), B and C subtilases, we demonstrated the enormous diversity of potentially useful detergent proteases in the recovered metagenomic DNA. Large- and small-insert metagenomic libraries were prepared in *Escherichia coli* and subjected to two different screening procedures. On one hand, sequence homology-based screening using above-mentioned degenerate primers was carried out. Since this approach, however, inherently limits the diversity of recovered proteases to subtilase-like enzymes, we also conducted an activity-based screening campaign. In total, 18 novel full-length sequences of diverse serine proteases were recovered, including three non-subtilase genes. Three enzymes – HP70, HP53 and HP23 – were found to be promising detergent enzymes with washing activities comparable to or even better than optimized *B. lentus* alkaline protease that was used as a benchmark.

Results and discussion

Evaluation of degenerate PCR primers for subtilases: a glimpse on metagenomic sequence diversity

Subtilase sequences share seven conserved regions (A–G, Fig. 1), three of them harbouring the catalytic triad residues Ser–His–Asp. Based on these regions, we derived CODEHOP primers (Rose *et al.*, 1998) targeting family A, B and C subtilases respectively. Additionally, a subset of family A, the alkaline subtilisins (family A-alk) was targeted separately, since these enzymes are often characterized by high activity and stability in the alkaline regime of washing processes. Eight primer pairs (Table 1) were found to yield distinct PCR products of about 550 bp on metagenomic DNA extracted from four different soils. Environmental samples were selected that have alkaline pH to increase the chances of finding exo-enzymes that are stable in the washing process. Sequencing efforts focused on two samples originating from a cement plant (S213 and S255, 48 sequenced clones). Comparable amounts of amplification product were obtained for all subtilase family primer pairs, suggesting particularly broad diversity in this sample. This may be due to the hostile, i.e. alkaline and nutrition depleted characteristics of this habitat that favour the enrichment of spore forming microorganisms generally containing one or several subtilase genes. Samples S256 and S247, in contrast, were biased towards family B and A/A-Alk PCR products, respectively, and only a limited amount of nine sequencing reactions was carried out for verification purposes. Interestingly, all 57 sequenced DNA stretches were not yet lodged in a database [30–83% identity to the nearest neighbour as identified by BLAST (Altschul *et al.*, 1997)] and unique – demonstrating the enormous sequence and, expectedly, functional diversity of subtilases in complex microbial habitats such as soil.

As shown in Fig. 2, the large majority of sequence tags classify into two major groups: group I, containing Proteinase K-type subtilases, and group II that includes the more closely related family A and B subtilases (Fig. 2). Only two out of 57 sequences formed a separate outgroup, demonstrating the target specificity of the used primer pairs. It should be noted, however, that a limited amount of inter-familiar cross-amplification was observed, particularly for family A and B primers, leading to further families of related enzymes within group I.

Recovery of subtilases from metagenomic libraries

Unfortunately, pre-evaluated cement plant DNA was too fragmented to allow the construction of a large-insert library. However, subtilase diversity seems to be generally high in soil samples, which is why we decided to initially screen an existing large-insert fosmid library prepared from

Table 1. Degenerate PCR primers for subtilase amplification.

Family	Name	Sequence	T_m (°C)	Degeneracy
A	NMP001	ACACATGTTGCCGGCacnrthgcngc	60.1	96
A	NMP005	AGCCACTCCAGCCACAtghggngwngc	60.9	96
A	NMP006	TTCCAAGATGATAATGGCcayggnacnca	60.3	32
A	NMP004	CACATGCGGTGTTgccatnswngt	62.7	64
A-alk	NMP041	CAATGGCGTTAAAGTTGCAgtnytngayac	60.2	64
A-alk	NMP042	gCCGACTTATCAAGATGGTAATggncayggnac	62.9	32
A-alk	NMP043	CCAGTGCTGCAACGccngcnacrtg	61.2	32
B	NMP028	CGTGGTGGTCGCCGTsrksswcacsg	61.4	128
B	NMP029	TGGCACGGCACCcackksgcsgg	62.6	16
B	NMP031	CGACCCAGGTGCCGtrgttsswgwa	60.2	32
B	NMP033	GACGCCGGCGACGTrsggsghsgcca	61.7	48
C	NMP034	CGTGATCTGCGGGccnggngcrwa	61.0	64
C	NMP037	CACCTGGGGCCTGGACmgnrtngayca	61.1	128

Two primer pairs were targeted at family A (NMP001/NMP005, NMP006/NMP004) and family A-alk (NMP041/NMP043 and NMP042/NMP043), three primer pairs were designed towards family B subtilases (NMP028/NMP031, NMP029/NMP031, NMP029/NMP033) and one primer pair amplified family C sequence tags (NMP037/NMP034).

Fig. 1. Target sites of degenerate subtilase PCR primers. An alignment of model sequences is shown to pinpoint the conserved regions (A–G, shaded in grey) that are targeted by the degenerate primers given in Table 1. Black boxes indicate family-specific sequence stretches targeted by respective primers (arrows). The catalytic triad (S–H–D) is comprised in regions B, C and G (labelled *). Phylogenetic analysis was conducted using tagged area as depicted (—).

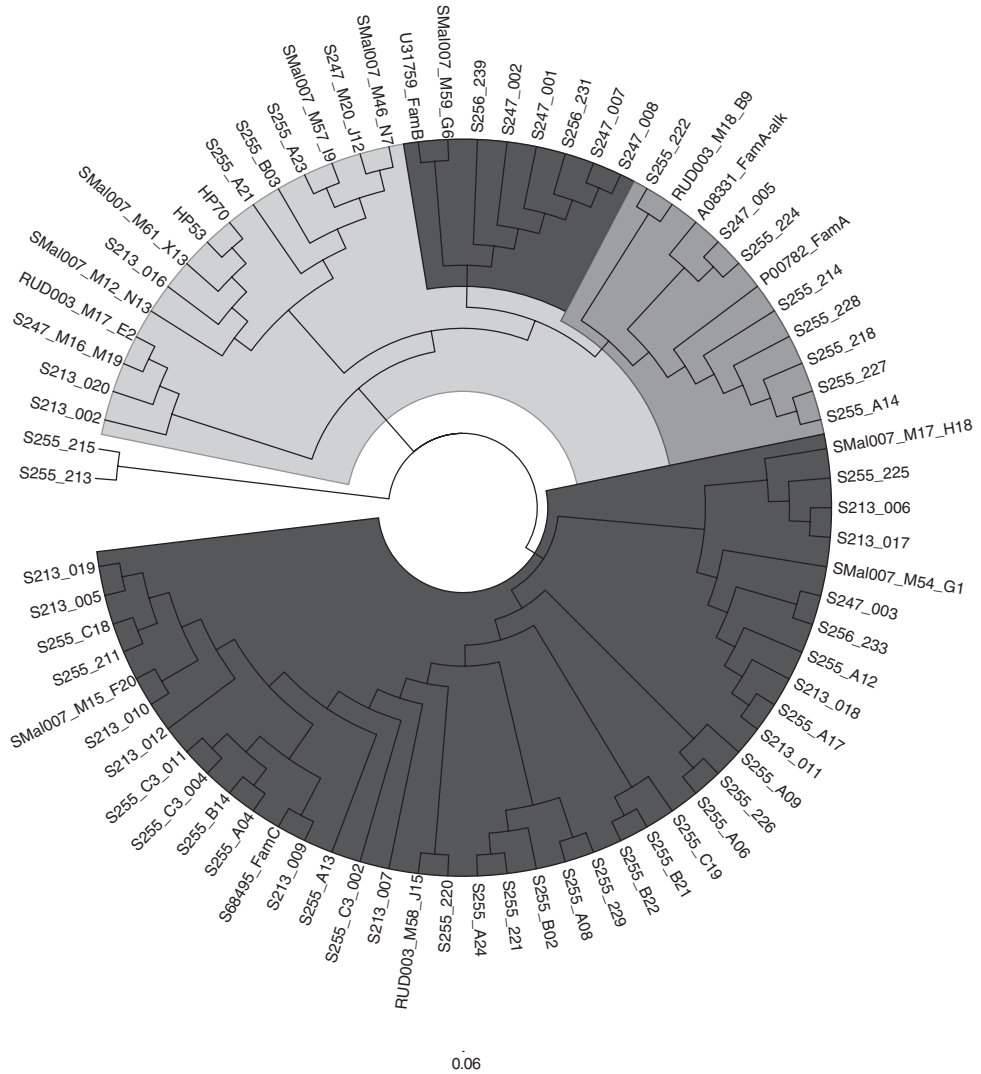

Fig. 2. Metagenomic subtilase sequence space. Metagenomic sequence tags as well as corresponding partial sequences of full-length subtilase genes recovered in this study (Table 3) have been used for the construction of a phylogenetic tree based on a CLUSTALW alignment (Jeanmougin *et al*., 1998). To pinpoint the location of the targeted subtilase families A, A-Alk, B and C, reference sequences have been incorporated into the tree (accession number followed by family name).

calcareous grassland soil DNA (RUD003) (Quaiser *et al*., 2002). Accordingly, another library was constructed from sample S247 that we though expected to be dominated by group I subtilases. To study gene dose effects on the success of screening and to countervail insert instabilities, another alkaline sample material (SMal) readily yielding large inserts was chosen for the construction of a library in a mid-copy cosmid vector. A further increase of expression level was realized in habitat DNA library HPXXIII by using a high-copy plasmid vector. In this case, DNA was prepared from soil microorganisms after a single cultivation step on solid medium to minimize experimental hurdles in DNA preparation. Since only a small fraction of microorganisms can be cultivated in this way, organisms from 12 different soil samples were pooled to increase diversity.

In total, large-insert libraries comprised 68 894 clones containing 2487 Mbp of metagenomic DNA (Table 2). Statistically, this corresponds to the nearly complete coverage of 127 microbial genomes (z) as calculated with a formula adapted from Clarke and Carbon [Clarke and Carbon, 1992): $z = l/G^*[1 - (1 - P)^{1/N}]$ with $N = 68\,894$, $P = 0.99$, $l = 34$ kb, and an average microbial genome size $G = 4000$ kb. A similar amount of genetic information was contained in the small-insert library HPXXIII (Table 2). To obtain an estimate of the number of subtilases to be found in our metagenomic libraries, we conducted a database search using the genomic information of publicly available prokaryotic genome sequences (http://cmr.jcvi.org/tigr-scripts/CMR/CmrHomePage.cgi; accessed on 10 December 2009). In 568 prokaryotic

Table 2. Metagenomic libraries constructed in this study.

Metagenome	Vector	No. of clones	Cloned DNA (Mbp)	Sequence-based hits	Activity-based hits
S247	EpiFos fosmid (single-copy)	11 520	461	2	0
RUD003	EpiFos fosmid (single-copy)	30 494	1220	3	0
SMal007	Expand III cosmid (mid-copy)	26 880	806	5	5
HPXXIII	pUC18 plasmid (high-copy)	309 375	2475	n.d.	3

n.d., not determined.

genome sequences, 299 subtilase sequences were found, suggesting that dozens of target genes could be present in our metagenomic libraries. However, it should be noted that genomes are evenly distributed in this artificial system, while in a real metagenome, abundant species not carrying target genes can decrease the actually observed hit frequencies. In fact, 18 clones carrying complete protease sequences were recovered by both activity- and sequence homology-based screening (Table 3). Most enzymes belonged to group I, including family A/A-Alk, B and related subtilases (Fig. 2). Notably, there was no overlap between enzymes discovered by the two different screening approaches. On one hand this is due to the fact that primary clones recovered by sequence homology-based screening did not show activity on skim milk agar. On the other hand primary clones identified by activity-based screening were not identified in the sequence homology-based screening campaign. We attribute this fact to biased pool DNA preparations used during high-throughput PCR screening. Namely, the expression of active target enzyme may cause severe toxic effects, leading to inhibited growth of host cells compared with inactive clones of the screened library and, consequently, to the under-representation of target DNA in the prepared DNA pools. These findings show that functional and homology-based screening systems are complementary when targeting novel subtilases and should be used in parallel.

Interestingly, only five target genes could be identified by activity-based screening, all originating from the cosmid microbial mat library. Apparently, expression levels in the single-copy fosmid libraries S247 and RUD003 were too low to allow detection. However, a further increase of expression level by using the high-copy plasmid vector pUC18 as a cloning vehicle did not boost the observed hit frequencies: only three active clones were found in

Table 3. Metagenomic protease clones.

Protease clone	ORF (bp)	MW (kDa)	Closest database homologue by BLASTX (Accession No.)	ID (%)	Signal (aa)[a]	Enzyme class	Accession No.
Sequence-based screening							
S247_M20_J12	1500	52.3	Protease, *Methylobacterium nodulans* (EDQ42311)	47	–	Subtilase	FN908460
S247_M16_M19	1833	64.1	Subtilisin, marine γ-proteobacterium (EAW39843)	37	26	Subtilase	FN908461
SMal007_M46_N7	1419	49	Protease, *M. nodulans* (EDQ42311)	58	–	Subtilase	FN908462
SMal007_M59_G6	2028	72	Alkaline serine protease, *Cyanothece* sp. (EAZ88556)	50	–	Subtilase	FN908463
SMal007_M61_X13	1794	61	Peptidase S8 and S53, *Dehalococcoides* sp. (EDO69483)	54	37	Subtilase	FN908464
SMal007_M12_N13	1875	63.5	Serine metalloprotease, *Burkholderia thailandensis* (ZP_02383698)	46	56	Subtilase	FN908465
SMal007_M15_F20	1182	40.3	Aqualysin-1, *Stigmatella aurantiaca* (EAU66137)	53	20	Subtilase	FN908466
RUD003_M17_E2	1263	42.1	Subtilisin, *Bacillus subtilis* (CAD62180)	37	25	Subtilase	FN908467
RUD003_M18_B9	1584	54.3	Subtilisin, *Geobacter sulfurreducens* (AAR35451)	40	28	Subtilase	FN908468
RUD003_M58_J15	1374	47.2	Extracellular serine proteinase, *S. aurantiaca* (EAU68410)	50	22	Subtilase	FN908469
Activity-based screening							
SMal007_M56_E12	1134	40.5	Peptidase, *Frankia* sp. (ABW14901)	47	–	S1 protease	FN908470
SMal007_M54_G6	978	34.5	Glutamyl endo peptidase, *Bacillus licheniformis* (AAU21945)	39	27	S1 protease	FN908471
SMal007_M57_I9	1305	45.4	Subtilisin, *Rhodococcus* sp. (ABG97200)	49	–	Subtilase	FN908472
SMal007_M17_H18	1701	60.7	Serine protease, *Streptomyces avermitilis* (NP_827825)	62	–	Subtilase	FN908473
SMal007_M54_G1	1554	54.3	Hypothetical protein, *Erythrobacter* sp. (EAQ30346)	66	64	Subtilase	FN908474
HP53	1761	58.5	Extracellular protease, *Stenotrophomonas maltophilia* (AAP13815)	85	38[b]	Subtilase	FN908475
HP70	1746	58.1	Extracellular protease, *S. maltophilia* (AAP13815)	84	32	Subtilase	FN908476
HP23	984	33.7	Hypothetical protein, *Arthrobacter chlorophenolicus* (EDS62663)	96	30	S1 protease	FN908477

a. aa, amino acids.
b. lacZα fusion.

2475 Mbp of cloned habitat DNA, which we attribute to the toxicity of highly expressed proteases.

Sequence analysis of metagenomic proteases

As expected in view of the screening strategy, all proteases isolated by the sequence homology-based screening approach classify as subtilases (S8 serine proteases), a group of enzymes that comprises all proteases commercially used in laundry and cleaning processes to date. With the activity-based screening approach, also three members of another group of enzymes (S1 serine proteases) were recovered besides subtilases (Table 3). S1 serine proteases, chymotrypsin being the most prominent member, are endopeptidases usually characterized by the presence of an N-terminal signal sequence and a propeptide that is cleaved off to attain activation of the enzyme. Interestingly, clone SMal007_M56_E12 encodes an active enzyme lacking such signal sequence.

According to the observed bias towards family A/A-Alk/B subtilases in total metagenomic DNA extracted from sample S247 (see above), also the two full-length genes recovered from the corresponding gene library classify as group I subtilases. Full-length genes belonging to the recovered sequence tags were not isolated, which may be attributed to restrictions of the applied screening methods or to the fact that metagenomic libraries usually only contain a fraction of the vast genetic information present in complex soil DNA. Compared with S247, protease sequences recovered from SMal and RUD003 show broader phylogenetic diversity with sequences belonging to both the major subtilase groups and different subgroups respectively (see Fig. 2).

All protease sequences were new, most of them displaying only between 37% and 66% amino acid sequence identity with the closest database entry. Only the sequences of HP53 and HP70 were 85% identical to a minor extracellular protease of *Stenotrophomonas maltophilia*. Remarkably, the latter two proteases share 97% of their amino acid residues, HP70 carrying a native 32-amino-acid signal sequence, while HP53 was found to be N-terminally fused to the lacZα peptide encoded by the pUC18 cloning vector. Regarding the high degree of homology between the two sequences, it can though be assumed that also the native HP53 protease possesses an N-terminal signal sequence. Although HP23 showed significant similarity to a hypothetical protein of *Arthrobacter chlorophenolicus*, it solely shared 52% of its amino acids with its closest verified protease homologue, a thermostable serine protease of the alkaliphilic and halophilic actinobacterium *Nesterenkonia* sp. (NAALP, Accession No. 3CP7_B).

Activity profiling of metagenomic subtilases

As shown above, a set of proteases with high sequence diversity was harvested from metagenomic sources. In order to evaluate their functional diversity, activity profiling was carried out in a microtitre plate format suited for high-throughput assaying of different parameters. As mentioned above, primary clones recovered by sequence homology-based screening did not show activity on skim milk agar, which may be due to the lack of expression signals suitable for *E. coli*. To ensure efficient transcription and initiation of translation, we cloned the protease sequences into a high-copy plasmid vector, providing a lacZα promoter as well as a suitable ribosome binding site. Two active clones were obtained in this way, SMal007_M61_X13_pUC and SMal007_M15_F20_pUC. The other eight pUC18 clones were viable but not active, indicating improper folding and secretion, respectively, or the absence of cofactors required for activity of the recombinant protease.

In view of problems related to heterologous gene expression, activity-based screening offers the inherent advantage of directly yielding active target clones. However, three of the clones recovered by this strategy were only moderately active and, therefore, were excluded from further biochemical characterization. Interestingly, one of these low-activity clones was HP70 that carries a native signal sequence, while its close relative HP53 displayed sufficient activity for further characterization.

In total, seven clones were selected for activity profiling, using BODIPY FL casein as a substrate, which gives rise to a fluorescence signal upon hydrolysis. Optimal pH and reaction temperatures were determined, as well as stability at high temperature (50°C) and high pH (10.9), and the influence of EDTA on enzyme activity. An expression clone of subtilisin Carlsberg (Jacobs *et al*., 1985) was used as a reference. Results are summarized in Table 4.

While all proteases showed a trend towards higher activity at elevated temperatures and high pH, activity profiles were quite diverse, particularly with respect to enzyme stability and tolerance towards EDTA. Subtilisin Carlsberg as well as SMal007_M57_I9 and HP23 were not inhibited by the presence of 1 mM EDTA. At 37°C, HP23 even showed an increase in activity, which is surprising since its sequence is similar to a number of S2-proteases that are known to be stabilized by divalent cations, particularly a metalloprotease of *Bacillus subtilis* [Mpr (Rufo *et al*., 1990)]. In contrast, a more or less pronounced decrease in activity was observed for the other proteases. Particularly HP53 seems to require the presence of divalent cations with only 10% (37°C) and 27% (50°C) residual activity in the presence of EDTA. With respect to stability, subtilisin Carlsberg, SMal007_M61_X13_pUC and SMal007_M57_I9 only

Table 4. Activity profiling of metagenomic serine proteases.

Clone	pH_{opt} at 37°C[a]	T_{opt} at pH 8.6 (°C)	Relative fluorescence (%) + 1 mM EDTA, pH 8.6, 37°C	Relative fluorescence (%) + 1 mM EDTA, pH 8.6, 50°C	Relative fluorescence (%) + pre-incubation,[b] pH 8.6, 37°C	Relative fluorescence (%) + pre-incubation,[b] pH 8.6, 50°C
Subtilisin Carlsberg	9.0	37, 50	101	100	89	85
SMal007_M61_X13_pUC	9.0	37	76	98	57	81
SMal007_M15_F20_pUC	8.6	50	90	66	7	10
SMal007_M56_E12	7.6–8.6	50	81	50	17	27
SMal007_M57_I9	8.6	50	101	93	82	86
SMal007_M17_H18	7.6–8.6	50	80	60	34	49
HP53	8.6	37	10	27	11	13
HP23	9.0	37	124	103	46	45

a. Identical results were obtained at 50°C.
b. Pre-incubation was carried out at pH 10.9 for 15 min at 50°C.

showed a slight inactivation upon pre-incubation at elevated temperature and pH. The other proteases were more affected, HP53 being most inactive with 11–13% remaining activity.

Laundry application tests

Interestingly, stability and EDTA inhibition data did not always correlate with the performance of proteases in a real, standardized small-scale laundry process. In fact, proteases HP23 and HP53 showed the highest potential for one of the envisaged applications, the use as a washing powder additive (Table 5). Both enzymes improved the removal of albuminous soiling at 40°C compared with a basic washing powder formulation, emphasizing their activity in presence of denaturing agents such as detergents and bleach. Towards blood/milk/ink soilings, protease HP53 even performed substantially better than the *B. lentus* alkaline protease that had been optimized for laundry purposes by site-directed mutagenesis (Christianson *et al*., 1995) and was used as a benchmark in our study. Due to the high sequence similarity to HP53, also HP70 may be a promising candidate for laundry applications. Indeed, expression optimization and biochemical characterization of an HP70 derivative have confirmed the high potential of this enzyme as a detergent protease (Ribitsch *et al*., 2010).

Experimental procedures

Environmental samples

Samples were collected from the upper layer (5–20 cm) of five sites having alkaline pH: grassland at Alsbach-Hähnlein, Germany (sample S247 pH 7.5), calcerous grassland at Darmstadt, Germany [sample RUD003, pH 8.0 (Quaiser *et al*., 2002)], a loess deposit at Bötzingen, Kaiserstuhl, Germany (sample S256, pH 8.5) and a cement plant at Gernsheim, Germany (samples S255, pH 8.5, and S213, pH 8.9). The cement plant samples were collected in close proximity to a limestone and a clay/iron ore dump respectively. Another sample was collected from microbial mats of an alkaline wastewater site in Zwingenberg (sample SMal007, pH 9.5). For the generation of a habitat sample (HPXXIII, see below), 12 soil samples were collected from the upper layer (5–20 cm) of different sites at the BRAIN Company campus in Zwingenberg. All of these samples had a neutral to slightly alkaline pH (7–7.4). Samples were stored at 4°C for up to 2 weeks before further processing.

Metagenomic and microbial DNA extraction

DNA samples serving as templates for PCR amplification were isolated using the FastDNA kit (BIO 101) that typically

Table 5. Laundry application tests of HP53 and HP23 at 40°C.

Basic washing powder additive	Soiling			
	A[a]	B[b]	C[c]	D[d]
None	20.1	18.7	n.d.	67.3
B. lentus alkaline protease, 2400 PE	25.5	27.9	n.d.	69.7
HP53, 2400 PE	48.1	60.7	n.d.	73.7
B. lentus alkaline protease, 6000 PE	32.6	41.1	n.d.	73.8
HP53, 6000 PE	42.9	53.5	n.d.	69.0
Standard deviation	4.5	1.7	–	2.4
None	21.8	14.4	50.4	n.d.
B. lentus alkaline protease, 2400 PE	29.3	24.2	70.3	n.d.
HP23, 2400 PE	23.8	16.6	61.7	n.d.
B. lentus alkaline protease, 6000 PE	33.9	33.0	71.4	n.d.
HP23, 6000 PE	26.6	19.2	67.9	n.d.
Standard deviation	1.4	1.5	2.3	–

a. Blood/milk/ink on cotton fabric.
b. Blood/milk/ink on a blended fabric of polyester and cotton.
c. Egg/grime on cotton fabric.
d. Blood on cotton.
Numbers indicate the degree of textile whiteness after washing as compared with barium sulfate that was defined as 100%.

yielded DNA fragments between 8 and 25 kb. Larger fragments for DNA library construction were obtained by using a direct lysis method adapted from Zhou and colleagues (1996). A habitat sample (HPXXIII) was generated by separately spreading soil suspensions of 12 Zwingenberg locations on HSP10 agar plates [per litre: 0.1 g of Yeast Extract (Difco), 0.1 g of Casein-Peptone (Difco), 0.1 g of soluble starch (Merck), 2 g of Na_2CO_3 (Sigma) and 50 g of Agar (Difco); pH 10]. Soil suspensions were obtained by dispersing 25 g of soil in 50 ml of 20 mM sodium phosphate buffer (pH 7.0) and homogenizing in a standard blender for three 1 min intervals with breaks in between to allow cooling. After settling of coarse particles, 200 μl were directly spread on round 10 cm HSP10 agar plates and incubated at 30°C for 2 weeks. Microbial cell lawns were washed from plates with sterile water and equal volumes of cell suspensions from each sample were combined. Cells collected by centrifugation were used for DNA extraction using the Geneclean Turbo Kit (Q-BIOgene).

Amplification of subtilase sequence tags

Degenerate primers (Table 1, Fig. 1) were designed using CODEHOP software (Rose *et al*., 1998). PCR amplifications were carried out using HotStarTaq Master Mix (Qiagen). Amplicons of the expected size (300–600 bp) were purified and directly cloned into the pCR2.1 cloning vector according to the manufacturer's instructions (Invitrogen). Inserts of at least three clones were sequenced.

Construction of metagenomic DNA libraries

Large-insert libraries. The environmental sample SMal007 served as a source of metagenomic DNA for the construction of a cosmid library, using the Expand Cloning kit (Roche). DNA fragments for cloning were obtained by subjecting metagenomic DNA preparations to pulsed-field gel electrophoresis (PFGE) and recovering 20–45 kb fragments with the QIAEX II gel extraction kit (Qiagen). Insert sizes were in the range of 25–36 kb. Large-insert libraries from soil samples RUD003 and S247 were constructed by use of the EpiFOS-Fosmid Library Production kit (Epicentre Biotechnologies). Insert sizes in this system were between 35 and 45 kb. Clones were arrayed in 384-well microtitre plates for screening purposes.

Small-insert library. A high-copy plasmid-based enzyme library was constructed from habitat sample HPXXIII. DNA fragments of 2–10 kb were obtained by partially digesting metagenomic DNA with AluI. After sizing by agarose gel electrophoresis and electroelution, fragments were ligated into SmaI-digested dephosphorylated pUC18 vector. All necessary enzymes were purchased from New England Biolabs (NEB) and used according to the instructions of the manufacturer. Ligation mixtures were used to transform *E. coli* DH12S (Gibco Life Technologies) by electroporation.

Activity-based screening

Metagenomic libraries were screened on skim milk agar medium that contained per litre 5 g of Yeast Extract (Difco), 10 g of Tryptone (Difco), 5 g of NaCl (Sigma), 20 g of Skim Milk (Difco) and 15 g of Agar (Difco) and antibiotic (25 μg ml^{-1} chloramphenicol and 100 μg ml^{-1} ampicillin respectively). Transformants were either directly spread on the solid medium (small-insert library) or arrayed large-insert libraries were replicated on the screening medium. After an overnight growth phase at 37°C, plates were incubated at 20–22°C for up to 4 weeks. Halo formation indicated the presence of clones with heterologous proteolytic activity. Full-length protease sequences of small-insert library hit clones were obtained by Sanger sequencing of the complete plasmid inserts. To avoid complete sequencing of 35–45 kb inserts, hit clones of large-insert libraries were subjected to transposon mutagenesis using the GPS-1 Genome Priming System (New England Biolabs). Sequences of mutants showing a protease-negative phenotype (no halo formation on skim milk agar) were elucidated by inside-out sequencing (Sanger method) starting from the known transposon sequence tag.

Sequence homology-based screening

Large-insert metagenomic gene libraries were replicated on LB agar containing 25 μg ml^{-1} chloramphenicol for fosmid libraries and 100 μg ml^{-1} ampicillin for cosmid libraries (1 microtitre plate per LB agar dish). After overnight incubation at 37°C, cells were washed from the LB agar dishes and episomal DNA was prepared from each dish using the QIAprep Spin Miniprep Kit (Qiagen) (pool DNA). For the first round of screening, three to five pool DNAs were combined and each of these 'super pool' DNAs was used as a template for parallel PCR amplifications with the degenerate subtilase primers listed in Table 1. Reaction conditions were as described above for the amplification of subtilase sequence tags. Super pools yielding a signal, i.e. a distinct PCR product of 300–600 bp, were subjected to a second round of screening. Here, successful PCR amplifications were repeated, using each of the three to five comprised pool DNAs separately as a template. A third round of screening was carried out using a colony filter hybridization technique to identify recombinant colonies causing PCR product formation in pool DNAs that tested positive. To obtain specific probes for the hybridization assay, PCR products obtained in the second round of screening were cloned into the pCR2.1 vector according to the manufacturer's instructions (Invitrogen). Inserts from three clones were sequenced and nested primers were designed to yield 200–250 bp products for each unique sequence tag. Using these primers, digoxigenin (DIG)-labelled probes were prepared by PCR using the PCR DIG Labelling Mix of Roche. Microtitre plates corresponding to the positive pool DNAs were replicated on LB agar and incubated overnight at 37°C. Colony filter hybridization was carried out according to the 'DIG Application Manual' (Roche Diagnostics). Chemoluminescent detection of the hybridized probe by anti-DIG-alkaline phosphatase Fab fragments and CPD-Star substrate was carried out according to the instructions of the manufacturer (Roche). Full-length sequences of hit clones were elucidated by a primer walking strategy and Sanger sequencing starting from the nested sequence tags. Sequences were submitted to GenBank under the accession numbers indicated in Table 3.

Activity profiling of metagenomic proteases

Activity assays were carried out using cultures of subtilase-expressing clones (host strain: *E. coli* DH10B) grown at 20°C and 180 r.p.m. rotary shaking for 20 days. To prepare crude cell extracts, cells were harvested by centrifugation, resuspended in 1/20 volume of ice-cold 50 mM phosphate buffer (pH 7) and disrupted by sonification. Supernatants were obtained by centrifugation at 5000 *g* for 5 min (4°C). Reactions were carried out in a total volume of 100 µl in microtitre plates (Opaque Plates, black; Corning BV Life Sciences), using BODIPY FL casein (Invitrogen) as a substrate (4.5 µg ml^{-1}). Reactions were buffered with 200 mM buffer according to the desired pH (Tris-HCl for pH 7.6 and 8.6 and Glycine-NaOH for pH 9.0). End-point measurements of fluorescence (λ_{ex} = 485 nm, λ_{em} = 520 nm) were carried out after 1 h of incubation at different temperatures and pH values, using a FLUOstar fluorescence reader (BMG Lab Technologies). Reactions were carried out in the presence of 1 mM EDTA at 37°C and 50°C at pH 8.6 to monitor the influence of divalent cations on enzyme activity. Stability under alkaline conditions and at elevated temperatures was tested by pre-incubating enzyme containing solutions for 15 min at pH 10.9 (adjusted with 1 M $NaCO_3$) and 50°C before carrying out the activity assay as described above. Reactions for all metagenomic proteases were carried out in parallel. Control reactions with crude cell extracts of *E. coli* DH10B carrying respective vectors without an insert were routinely performed and did not show any detectable background activity.

Laundry application assays

Textiles with standardized soiling were obtained from the Swiss Federal Laboratories for Materials Testing and Research (EMPA). Textiles were washed for 30 min at 40°C, using 5.9 g of washing powder per litre of water (16° German water hardness). The basic washing powder contained 4% (w/v) linear alkyl benzene sulfonate (sodium salt), 4% (w/v) C_{12}–C_{18} fatty alcohol sulfate (sodium salt), 5.5% (w/v) C_{12}–C_{18} fatty alcohol with 7EO, 1% (w/v) sodium soap, 11% (w/v) sodium carbonate, 2.5% (w/v) amorphous sodium bisilicate, 20% (w/v) sodium perborate tetrahydrate, 5.5% (w/v) tetra-acetyl ethylene diamine (TAED), 25% zeolite A, 4.5% (w/v) polycarboxylate, 0.5% (w/v) phosphonate, 2.5% (w/v) granulated foaming inhibitor, 5% (w/v) sodium sulfate, rest: water, optical bleach, salts. Basic washing powder without further additives was used as a negative control. As a positive control, basic washing powder was supplemented with 0.2% (w/v) and 0.5% (w/v) *B. lentus* alkaline protease, yielding a specific activity of 2400 and 6000 PE [protease units as determined according to van Raay and colleagues (1970)], respectively, per litre of washing liquor. Corresponding activity concentrations were used for the proteases identified in this study. After the washing, the degree of whiteness of the washed textiles was compared with the one of barium sulfate that was defined as 100%. Triplicate measurements were carried out using a Datacolor SF500-2 spectrometer (Datacolor) at 460 nm, 30 mm aperture, no brilliance, illuminant D65, 10°, d/8°.

References

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25:** 3389–3402.

Amann, R.I., Ludwig, W., and Schleifer, K.H. (1995) Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiol Rev* **59:** 143–169.

Christianson, T.M., Goddette, D.W., Kottwitz, B., Maurer, K.-H., Paech, C., Pochandke, W., *et al*. (1995) Improved enzymes and detergents containing them. WO 95/23221 A1.

Clarke, L., and Carbon, J. (1992) A colony bank containing synthetic Col El hybrid plasmids representative of the entire *E. coli* genome. 1976. *Biotechnology* **24:** 179–187.

Eck, J., Gabor, E., Liebeton, K., Meurer, G., and Niehaus, F. (2009) From prospecting to product: industrial metagenomics is coming of age. In *Protein Engineering Handbook*. Lutz, S., and Bornscheuer, U.T. (eds). Weinheim, Germany: Wiley-VCH, pp. 295–324.

Goddette, D.W., Paech, C., Yang, S.S., Mielenz, J.R., Bystroff, C., Wilke, M.E., and Fletterick, R.J. (1992) The crystal structure of the *Bacillus lentus* alkaline protease, subtilisin BL, at 1.4 A resolution. *J Mol Biol* **228:** 580–595.

Handelsman, J., Rondon, M.R., Brady, S.F., Clardy, J., and Goodman, R.M. (1998) Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem Biol* **5:** R245–R249.

Herbots, I. (2007) Enzymes in household detergents. In *Enzymes in Industry*. Ähle, W. (ed.). Weinheim, Germany: Wiley-VCH, pp. 154–178.

Jacobs, M., Eliasson, M., Uhlen, M., and Flock, J.I. (1985) Cloning, sequencing and expression of subtilisin Carlsberg from *Bacillus licheniformis*. *Nucleic Acids Res* **13:** 8913–8926.

Jeanmougin, F., Thompson, J., Gouy, M., Higgins, D.G., and Gibson, T.J. (1998) Multiple sequence alignment with Clustal X. *Trends Biochem Sci* **23:** 403–405.

Quaiser, A., Ochsenreiter, T., Klenk, H.P., Kletzin, A., Treusch, A.H., Meurer, G., *et al*. (2002) First insight into the genome of an uncultivated crenarchaeote from soil. *Environ Microbiol* **4:** 603–611.

van Raay, H.G., Saran, H., and Verbeek, H. (1970) Zur Bestimmung der proteolylischen Aktivitat in Enzymkonzentraten und enzymhaltigen Wasch-, Spül- und Reinigungsmitteln. *Tenside* **7:** 125–132.

Ribitsch, D., Karl, W., Birner-Gruenberger, R., Gruber, K., Eiteljoerg, I., Remler, P., *et al*. (2010) C-terminal truncation of a metagenome-derived detergent protease for effective expression in *E. coli*. *J Biotechnol* **150:** 408–416.

Rose, T.M., Schultz, E.R., Henikoff, J.G., Pietrokovski, S., McCallum, C.M., and Henikoff, S. (1998) Consensus-degenerate hybrid oligonucleotide primers for amplificationof distantly related sequences. *Nucleic Acids Res* **26:** 1628–1635.

Rufo, G.A., Jr, Sullivan, B.J., Sloma, A., and Pero, J. (1990) Isolation and characterization of a novel extracellular

metalloprotease from *Bacillus subtilis*. *J Bacteriol* **172:** 1019–1023.

Siezen, R.J., and Leunissen, J.A. (1997) Subtilases: the superfamily of subtilisin-like serine proteases. *Protein Sci* **6:** 501–523.

Smith, E.L., DeLange, R.J., Evans, W.H., Landon, M., and Markland, F.S. (1968) Subtilisin Carlsberg. V. The complete sequence; comparison with subtilisin BPN'; evolutionary relationships. *J Biol Chem* **243:** 2184–2191.

Vasantha, N., Thompson, L.D., Rhodes, C., Banner, C., Nagle, J., and Filpula, D. (1984) Genes for alkaline protease and neutral protease from *Bacillus amyloliquefaciens* contain a large open reading frame between the regions coding for signal sequence and mature protein. *J Bacteriol* **159:** 811–819.

Wells, J.A., Ferrari, E., Henner, D.J., Estell, D.A., and Chen, E.Y. (1983) Cloning, sequencing, and secretion of *Bacillus amyloliquefaciens* subtilisin in *Bacillus subtilis*. *Nucleic Acids Res* **11:** 7911–7925.

Zhou, J., Bruns, M.A., and Tiedje, J.M. (1996) DNA recovery from soils of diverse composition. *Appl Environ Microbiol* **62:** 316–322.