Environmental Microbiology (2004) doi:10.1111/j.1462-2920.2004.00643.x

Construction, characterization, and use of small-insert gene banks of DNA isolated from soil and enrichment cultures for the recovery of novel amidases

Esther M. Gabor, Erik J. de Vries and Dick B. Janssen*
Department of Biochemistry, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Nijenborgh 4, 9747 AG Groningen, the Netherlands.

Summary

To obtain new amidases of biocatalytic relevance, we used microorganisms indigenous to different types of soil and sediment as a source of DNA for the construction of environmental gene banks, following two different strategies. In one case, DNA was isolated from soil without preceding cultivation to preserve a high degree of (phylo)genetic diversity. Alternatively, DNA samples were obtained from enrichment cultures, which is thought to reduce the number of clones required to find a target enzyme. To selectively sustain the growth of organisms exhibiting amidase activity, cultures were supplied with a single amide or a mixture of different aromatic and non-aromatic acetamide and glycine amide derivatives as the only nitrogen source. Metagenomic DNA was cloned into a high-copy plasmid vector and transferred to *E. coli*, and the resulting gene banks were searched for positives by growth selection. In this way, we isolated a number of recombinant *E. coli* strains with a stable phenotype, each expressing an amidase with a distinct substrate profile. One of these clones was found to produce a new and highly active penicillin amidase, a promising biocatalyst that may allow higher yields in the enzymatic synthesis of β-lactam antibiotics.

Introduction

A major portion of microorganisms present in nature cannot readily be obtained in pure culture, which is why culture-independent techniques become more and more important for assessing diversity and exploiting the biocatalytic potential of microbial communities. The construction and screening of gene libraries prepared from DNA directly isolated from environmental samples is a recent and powerful tool for the discovery of new enzymes of biotechnological interest [e.g. lipases and esterases (Henne *et al*., 2000), chitinases (Cottrell *et al*., 1999), cellulases (Healy *et al*., 1995), polyketide synthases (Courtois *et al*., 2003), and amylases (Rondon *et al*., 2000); for review see (Lorenz *et al*., 2002)]. Whereas standard methods based on the screening of isolated microorganisms are inherently limited to the tiny fraction of cultivable microbial species [<1% (Torsvik *et al*., 2002)], environmental gene banks in principle provide access to the entire sequence space present in nature (Handelsman *et al*., 1998).

One of the challenges of environmental cloning is the immense number of transformants that needs to be produced and screened. It has been estimated that more than 10^7 plasmid clones (5 kb inserts) or 10^6 BAC clones (100 kb inserts) would be required in order to represent the collective genomes, i.e. the metagenome (Handelsman *et al*., 1998), of several thousands different species as typically present in a soil sample, assuming the idealized case of all species being equally abundant (Lorenz and Schleper, 2002). As a prerequisite to reach these large numbers, efficient and non-biased methods need to be employed for the extraction and purification of high-quality DNA from environmental samples (Gabor *et al*., 2003). The subsequent construction of gene banks from environmental DNA involves the same steps and techniques as the cloning of genomes of single microorganisms, i.e. fragmentation of genomic DNA, ligation into an appropriate vector system, and transformation to a host organism, usually *E. coli* (Daniel, 2002). However, contaminants such as humic acids in the starting material, which are always co-extracted with nucleic acids from environmental samples like soil, affect enzymatic steps in the procedure (digestion of DNA and ligation) and decrease transformation efficiencies (Tebbe and Vahjen, 1993). Regarding the large numbers of clones that need to be produced, extreme care must therefore be taken to optimize each step in the used protocol.

In an ideal gene library with equal representation of all indigenous species and no non-productive clones, the number of clones that statistically need to be

Received 7 January, 2004; accepted 30 March, 2004. *For correspondence. E-mail D.B.Janssen@chem.rug.nl; Tel. (+31) 50 3634209; Fax (+31) 50 3634165.

screened to find a positive one is solely determined by the frequency of organisms carrying one or several genes of interest in the source DNA. This frequency can be raised by a classical enrichment step preceding DNA isolation, where organisms are cultivated under selective pressure that favours the growth of bacteria expressing a desired activity. Like traditional strategies for enzyme discovery, this method may suffer from the fact that many organisms will not grow under laboratory conditions because of their special requirements in (nutritional) growth conditions. Many so-called 'non-cultivable' bacteria, however, cannot be obtained as pure isolates because they exist in nature as symbionts or as part of consortia, requiring the presence of certain other organisms for growth (Schink, 2002). Those organisms may be maintained in mixed liquid culture, making enrichment cultures a potential source of new genes from organisms that are not accessible by traditional techniques. Besides the anticipated decrease in gene bank size, the use of enrichment cultures as a starting material for cloning also facilitates DNA recovery and handling, thereby allowing faster access to the environmental gene pool. So far, this strategy has yielded new biotin biosynthesis operons (Entcheva *et al*., 2000) and genes conferring alcohol oxidoreductase activity (Knietsch *et al*., 2003).

In this paper, we present our results from both direct environmental cloning and cloning from enrichment cultures for the isolation of new amidases, using growth selection of active clones on solid agar media. With these approaches, six clones comprising recombinant DNA with no or only low homology to existing database-entries were isolated. One of the positives expressed a novel penicillin amidase (penicillin acylase, EC 3.5.1.11) with highly favourable enzymatic properties, which demonstrates the potential of the applied approach to yield industrially relevant biocatalysts.

Results and discussion

Diversity of source DNA

In this study, we were interested in the isolation of new enzymes that can convert derivatives of D-phenylglycine amide. These molecules are important intermediates in the biocatalytic synthesis of semisynthetic β-lactam antibiotics, of which ampicillin, amoxicillin, cephalexin and cefadroxil are prominent examples (Bruggink *et al*., 1998). To increase the chance of finding useful enzymes, we enriched marine sludge, sand and loamy soil organisms on D-phenylglycine amide as sole nitrogen source, selecting organisms that can hydrolyse this compound and release ammonia (enrichments ENR-M, ENR-S, and ENR-L; Fig. 1A). Glucose, glycerol and sodium acetate were supplied as easily degradable carbon sources. To prevent the cultivation of only the fastest growing species and preserve a high degree of diversity of target organisms, the cultivation time was kept short in all experiments with only one transfer to fresh medium. As a measure to sustain diversity also in substrate specificity, a mixture of 10 different amides (Fig. 1B, compounds 1–10) was used for the cultivation of goose pond microorganisms (ENR-G). All enrichments were carried out under nitrogen-limiting conditions, which was confirmed by the absence of growth in parallel cultures to which no nitrogen source was added. Consequently, the selection of nitrogen-fixing organisms under the applied conditions can also be excluded.

To estimate the bacterial diversity in the different cultures, 16S rRNA fragments were PCR-amplified from the isolated DNA and separated by denaturing gradient gel electrophoresis (DGGE) (Fig. 2). Depending on the inocula, 10–16 different bands were observed, indicating the presence of relatively low numbers of abundant organisms. In contrast, a much higher phylogenetic diversity (44 bands) was detected when analysing metagenomic DNA directly extracted from the same loam soil sample used as the inoculum for enrichment ENR-L (Gabor *et al*., 2003). This shows that a major portion of bacterial species was lost during enrichment, which may be due to the inability of certain species to metabolize the provided nitrogen source (intended effect) or due to the applied general growth conditions (adverse effect), e.g. cultivation in liquid medium, supplied growth substrates, temperature, aeration, or competition between numerically inferior and superior bacterial populations (Santegoeds *et al*., 1996). To distinguish between these two situations, PCR-DGGE analysis was also carried out for a control culture that was set up in the same way as ENR-G, except that no selective nitrogen source but readily accessible NH_4Cl was added. Figure 2 (lanes 4 and 5) reveals a similar degree of diversity with clear differences concerning the location and intensity of bands in the two DGGE patterns, which suggests that most of the enriched organisms were able to grow because of their capacity to (co)metabolize the selective compounds. This conclusion is supported by an almost complete (>90%) depletion of amides 1–4 and 7–9 (Fig. 1B) in ENR-G, as determined by HPLC analysis of the culture medium before inoculation and after cell collection. The chemical stability of the supplied amides under the assay conditions was such that no spontaneous hydrolysis occurred. The degradation of D-mandelamide, α-methylphenylglycine amide, and α-methylvaline could not be monitored because of overlapping substrate and product peaks. Complete degradation of the supplied nitrogen source phenylglycine amide was also observed for ENR-M, ENR-S, and ENR-L.

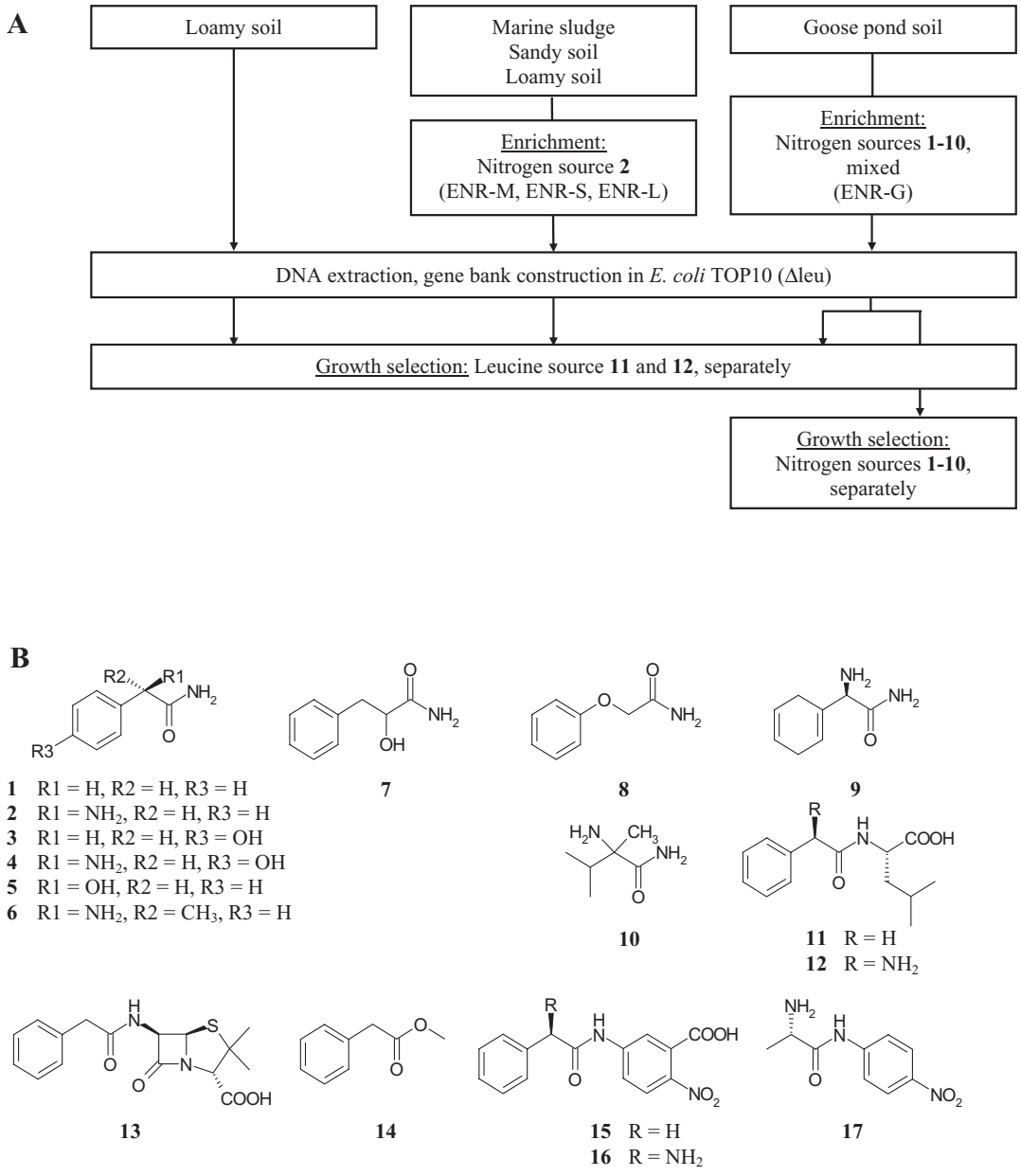

Fig. 1. A. Strategy for the isolation of amidase-expressing clones followed in this study.
B. Substrates used for enrichment (1–10), selection media (1–12), and substrate profiling (2, 3, 5, 8, 9, 11 and 13–17). Phenylacetamide (1), D-phenylglycine amide (2), *p*-hydroxyphenylacetamide (3), *p*-hydroxy-D-phenylglycine amide (4), D-mandelamide (5), α-methylphenylglycine amide (6, racemic mixture), α-hydroxy-3-phenyl-propionamide (7, racemic mixture), phenoxyacetamide (8), D-dihydrophenylglycine amide (9), α-methylvaline (10, racemic mixture), phenylacetyl-L-leucine (11), D-phenylglycine-L-leucine (12), penicillin G (13), phenylacetic acid methyl ester (14), 2-nitro-5-[(phenylacetyl)amino]-benzoic acid (NIPAB, 15), D-2-nitro-5-[(phenylglycyl)amino]-benzoic acid (NIPGB, 16), alanine-*p*-nitroanilide (APNA, 17). While the conversion of compounds 1–14 was followed by HPLC, hydrolysis of the colorimetric compounds (15–17) could be observed photometrically at 405 nm.

Gene bank construction

For DNA extraction from enrichment cultures, we used an optimized protocol that allowed the parallel processing of all four samples and yielded large amounts of high molecular weight DNA with no need for further purification. Even organisms that are known to be difficult to lyse like *Rhodococcus* or *Mycobacterium* sp. were readily accessible as established in preliminary experiments. Enrichment DNA and nucleic acids isolated from loamy soil without preceding cultivation (Gabor *et al.*, 2003) was used to construct five different gene banks (Table 1).

In a first step, high molecular weight DNA was mechanically fragmented using a nebulizer (Bodenteich *et al.*,

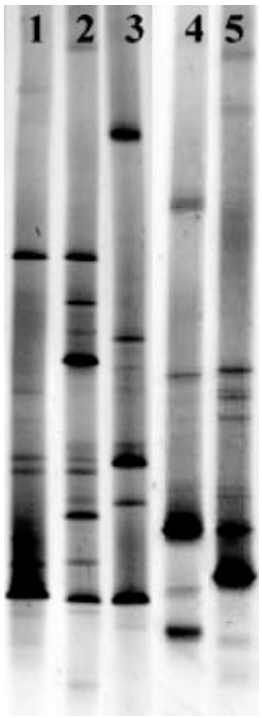

Fig. 2. DGGE profiles of DNA isolated from enrichment cultures. Lane 1, ENR-M; lane 2, ENR-S; lane 3, ENR-L; lane 4, control culture for ENR-G (1 mM NH_4Cl added as sole source of nitrogen); lane 5, ENR-G.

1994). In this small plastic device, DNA solutions are squeezed through small pores by overpressure, creating shearing forces that cause DNA strands to break. By adjusting pressure and shearing time, we were able to readily produce fragments between 1 and 10 kb at will with size distributions that were more narrow than those obtained in enzymatic digests (data not shown), which allows the efficient use of a possibly limited amount of source DNA. A further advantage of mechanical compared with enzymatic fragmentation is the fact that it is truly random and not dependent on restriction sites or methylation patterns, which may bias genome representation in the produced gene bank (Oefner *et al*., 1996). Furthermore, enzymatic restriction may be inhibited by contaminants in the DNA extract, which can especially be a problem when working with nucleic acids directly isolated from environmental samples. Fragments were size-selected on gel, blunted with Klenow and T4 polymerase, and ligated into the *Eco*RV site of plasmid pZero-2, a high-copy number vector. A lethal gene (*ccdB*) that is disrupted and consequently not expressed when an insert is incorporated allowed the construction of gene banks with background levels of self-ligated vector of ≤5%. Contaminants present in the DNA extract prepared from loamy soil could not be entirely removed during gel purification, which only allowed the production of about 30% of the number of clones that was produced from a corresponding amount of the more pure enrichment DNA. In average, gene banks contained inserts of 5.2 kb.

Estimation of the minimally required gene bank size

The number of clones that needs to be produced to cover the collective genomes present in a DNA extract can be estimated by means of DGGE patterns. This is particularly true when working with DNA isolated from an enrichment culture that is of solely prokaryotic origin. DNA extracts obtained from various environmental samples by direct lysis methods, in contrast, have been shown to contain major portions (>60%) of eukaryotic DNA (Gabor *et al*., 2003), leading to an increase in the required library size due to the formation of non-productive clones with eukaryotic inserts, virtually diluting the desired prokaryotic recombinants. Although this number can often be limited by the use of cell extraction-based DNA isolation protocols, a certain fraction of eukaryotic DNA is always co-extracted.

Based on the assumption that expression of environmental genes is mostly dependent on native promoters (Daniel, 2002) and that bacterial species are equally abundant in the DNA extract, the minimal gene bank size N_{min} can be calculated by $N_{min} = \ln(1\text{-}P)\ \ln/(1\text{-}(I\text{-}X)/$

Table 1. Diversity of source DNA and gene bank size

DNA source[a]	No. of DGGE bands, *n*	Minimal gene bank size[b], *N*	Actual gene bank size	No. of positives (plasmid name)
ENR-M	11	18 000 (31 000)	23 000	1 (pM1)
ENR-S	14	23 000 (40 000)	30 000	2 (pS1 and pS2)
ENR-L	10	17 000 (29 000)	35 000	0
ENR-G	16	27 000 (46 000)	25 000	2 (pG1 and pG2)
Loamy soil	44	60 000 (126 000)	80 000	2 (pL1 and pL2)

a. As inocula for enrichments, marine sludge (ENR-M), sand soil (ENR-S), loamy soil from an agricultural field (ENR-L), and soil collected at a goose pond (ENR-G) were used. From the loamy soil, DNA was also extracted without preceding cultivation (last row).
b. The number of minimally required clones was calculated by Equation 1 with $P = 0.9$, $I = 5.2$ kb, $X = 0.9$ kb, and $G = 3100$ kb, assuming gene expression from native promoters and even distribution of the different genomes. The numbers in brackets were calculated with $X = 2.7$ kb, which is a typical gene size of penicillin amidases.

($n{\cdot}G$)) (Equation 1). Here, P is the probability to find each environmental gene at least once, n is the number of different species present, I is the average insert size in the gene bank, X is the average gene size, and G is the average genome size. Whereas X can be neglected when the average insert size is much bigger (e.g. in BAC libraries), it needs to be taken into account for small insert gene banks as constructed in this study. In view of most microorganisms not being discovered yet, parameters X and G can only be roughly estimated on basis of already sequenced and annotated organisms. When analysing the chromosome sequences of all 95 completely sequenced prokaryotic organisms currently available at the TIGR database, neglecting alternative species of the same genus, G was found to be 3100 kb. The average length of genes in the database (X) is 0.9 kb. While P can be chosen and I is adjusted in the procedure at will, n needs to be determined experimentally.

In a DGGE profile, distinct bands correspond to 16S rRNA fragments of different melting behaviour resulting from sequence differences, which are in turn caused by a different phylogenetic origin. Consequently, individual PCR fragments represent unique bacterial populations (and possibly different 16S rRNA genes within a single species) and can be used as a rough estimate for n. It needs to be taken into account that PCR-DGGE only reveals populations that account for more than 0.1% of the total community (Gelsomino *et al*., 1999), thereby failing to detect the majority of bacterial species present in, e.g. soil samples. However, as organisms of lower abundance are unlikely to be significantly represented in small insert gene banks as typically constructed nowadays ($\leq 10^6$ clones), DGGE profiles still provide useful information on the number of different genomes that will constitute the major part of a resulting gene bank.

As shown in Table 1, all enrichment gene banks were (nearly) complete with respect to the calculated minimal gene bank sizes and allowed the isolation of 1 or 2 clones exhibiting the desired activity. In principle, one would expect to recover at least one positive clone per species that passed the selection procedure (i.e. 10–16), provided that a suitable screening method is used. However, expression of the active protein may fail due to the lack of proper expression signals, incorrect folding or processing, nonavailability of crucial cofactors, etc. Furthermore, some of the detected organisms may have only been able to withstand the selective pressure by using nitrogen-containing compounds or excess ammonia released by the primary degraders of the provided amides, not containing any active amidase themselves. Another reason for the apparent discrepancy, however, can be found in the calculation of N_{min} itself, which is based on the idealized assumption that all species are equally abundant in the source DNA. As can be seen from the DGGE profiles, band intensities though vary by more than a factor 10, meaning that N_{min} considerably underestimates the number of clones that is required to represent the collective genomes. Another parameter that strongly influences the outcome of Equation 1 is X. Whereas the average gene size in microorganisms is relatively small, specific classes of coding sequences can be much larger. Penicillin acylases, our primary targets, for instance, are typically three times larger, leading to minimal library sizes that are significantly bigger as a result of the decreased chance to clone a complete gene on DNA inserts of a given size (Table 1, in brackets). Therefore, more recombinants would need to be screened for a comprehensive mining of the biocatalytic potential of organisms present in the enrichment cultures.

Recovery of clones with amidase activity

Clones exhibiting amidase activity were selected on a medium containing phenylacetyl-L-leucine or D-phenylglycine-L-leucine (Fig. 1B, compounds 11 and 12) as the sole source of leucine, allowing only the growth of recombinants that can complement the leucine auxotrophy of the *E. coli* TOP10 host strain by hydrolysing the selective amide (Forney and Wong, 1989). As those compounds resemble D-phenylglycine amide, the selective nitrogen source used for cultivation of marine sludge, sand, and loam soil organisms, we expected to recover most of the amidases that conferred the ability to grow to their host organisms. The loamy soil gene bank produced without cultivation was screened accordingly.

For enrichment culture ENR-G, a mixture of different N-sources was employed, which possibly sustained the growth of organisms carrying enzymes with substrate specificities other than for phenylacetylated or D-phenylglycylated compounds. Besides leucine-limited medium, we therefore used nitrogen-limited agar media, in which amides 1–10 (Fig. 1) were separately supplied as the sole nitrogen source. In this way, two clones were found in gene bank G, one growing the best on D-phenylglycine amide-containing medium (plasmid pG1) and the other on D-mandelamide-containing medium (plasmid pG2). Activity of the latter clone, however, was lost when it was transferred to fresh medium, leading to its exclusion from further analysis. From the other gene banks, four unique clones with a stable amidase phenotype were isolated on phenylacetyl-L-leucine, and one (plasmid pS1) was found on the Cα-amino derivative (Table 1). To assure that only plasmid-encoded activity was selected, recombinant plasmids were isolated, re-transformed to *E. coli* TOP10, and plated on the selective medium on which they were originally isolated.

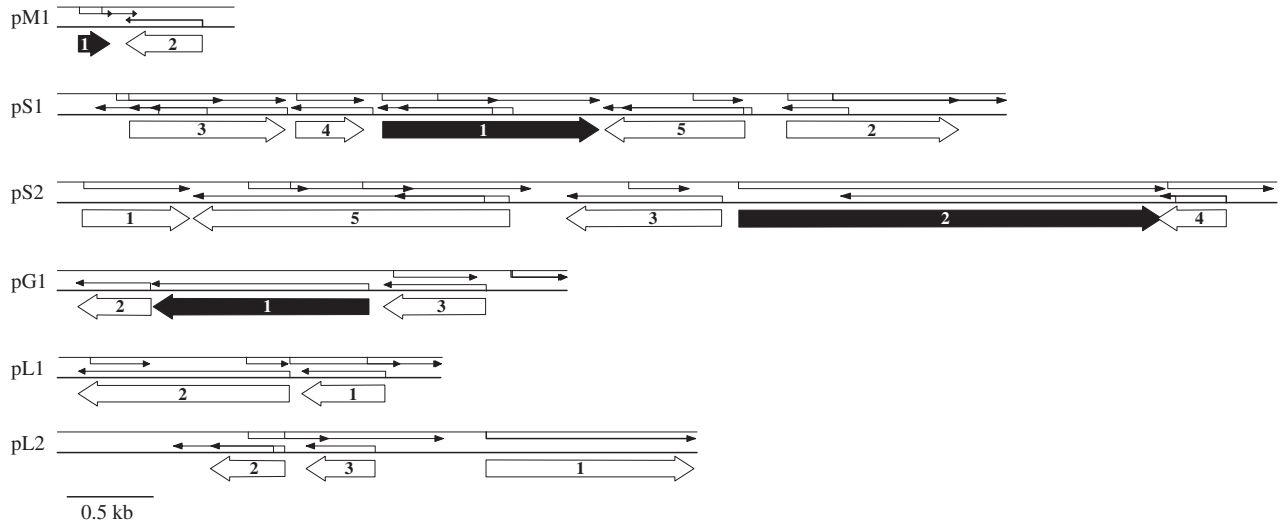

Fig. 3. Genetic organization of environmental insert DNA encoding enzymes with amidase activity. The promoter of the vector (P_{lac}) is located on the left side (upper strand) of the shown inserts. Black arrows indicate ORFs that are responsible for the observed enzyme activity. ORF numbers of each plasmid refer to those given in Table 2.

Sequence analysis of the recombinants

The six amidase activity-exhibiting clones were shown to contain a total of 50 heterologous ORFs with the major part having no homology to any GenBank database entry (Fig. 3). Only 20 ORFs yielded database hits with moderate (25%) to high (91%) identity scores, including seven predicted coding sequences of unknown function (Table 2). No single hit revealed complete identity with a known protein, which illustrates the potential of the approach taken for accessing unexplored sequence space. In four cases, the ORF responsible for the

Table 2. Amidase activity encoding plasmids, identified ORFs, and sequence similarities.

Plasmid (insert size, total no. of ORFs[a])	ORF[b] (no. of encoded amino acids)	Closest database homologue (accession number)	Organism	Identity[c]
pM1 (1.017 kb, 3)	**ORF1** (92)	**putative hydrolase** (NP_461274)	*Salmonella typhimurium*	91% (67/199)
	ORF2 (243)	putative transport protein (NP_416795)	*Escherichia coli*	87% (233/610)*
pS1 (5.461 kb, 17)	**ORF1** (443)	**probable Xaa-Pro dipeptidase** (NP_419119)	*Caulobacter crescentus*	41% (427/428)*
	ORF2 (348)	TonB-dependent receptor (NP_420476)	*Caulobacter crescentus*	58% (331/809)*
	ORF3 (318)	SalR transcriptional repressor (NP_639440)	*Xanthomonas campestris*	37% (282/361)
	ORF4 (136)	hypothetical protein (B87530)	*Caulobacter crescentus*	65% (96/101)*
	ORF5 (285)	hypothetical protein (SCE20–22)	*Streptomyces coelicolor*	46% (282/290)
pS2 (7.019 kb, 12)	ORF1 (215)	hypothetical protein (ZP_00118494)	*Magnetospirillum magnetotacticum*	46% (178/287)
	ORF2 (874)	**penicillin G amidase** (AAP20806)	*Achromobacter xylosoxidans*	84% (822/843)*
	ORF3 (319)	transcriptional activation protein (P52686)	*Pseudomonas* sp. ATCC19151	31% (282/306)
	ORF4 (135)	hypothetical protein (ZP_00050555)	*Cytophaga hutchinsonii*	63% (110/612)
	ORF5 (648)	ABC transporter protein (NP_437604.1)	*Sinorhizobium meliloti*	52% (602/604)*
pG1 (2.933 kb, 6)	**ORF1** (495)	**mandelamide hydrolase** (AAO23019)	*Pseudomonas putida*	68% (466/507)
	ORF2 (168)	benzaldehyde dehydrogenase (AAO23020)	*Pseudomonas putida*	79% (143/436)*
	ORF3 (233)	putative regulatory protein (AAO23018)	*Pseudomonas putida*	63% (149/233)
pL1 (2.216 kb, 6)	ORF1 (203)	probable peroxidase (G64843)	*Escherichia coli*	33% (190/243)*
	ORF2 (512)	hypothetical protein (NP_661777)	*Chlorobium tepidum*	29% (454/522)
pL2 (3.681 kb, 6)	ORF1 (451)	hypothetical protein (NP_772653)	*Bradyrhizobium japonicum*	25% (271/1195)*
	ORF2 (152)	hypothetical protein (NP_442421)	*Synechocystis* sp.	37% (114/148)
	ORF3 (141)	anthranilate phosphoribosyltransferase (NP_718589)	*Shewanella oneidensis*	40% (76/347)*

a. ORFs with alternative start sites for the same gene were counted as one and are represented by the largest ORF.
b. ORFs that are not listed did not show significant homology to any database entry. ORFs marked in bold were identified to be responsible for the observed enzyme activity.
c. The numbers in brackets indicate the length of the homologous amino acid sequence stretch with respect to the protein length of the database hit. Protein sequences derived of ORFs that are marked by a star (*) were predicted to carry a signal sequence, using SignalP software (Henrik *et al.*, 1997).

Table 3. Relative activities of the six recovered environmental clones on different substrates.

Compound[a]	Relative activity of clones carrying a recombinant plasmid [%][b]					
	pM1	pS1	pS2	pG1	pL1	pL2
D-Phenylglycine amide (2)	10	–	56	100	–	38
p-Hydroxy-D-phenylglycine amide (3)	33	–	65	37	–	52
D-Mandelamide (5)	–	–	46	33	–	–
Phenoxyacetamide (8)	–	–	4	34	–	34
D-Dihydrophenylglycine amide (9)	–	–	–	7	–	–
Phenylacetyl-L-leucine (11)	82	100	79	7	20	45
Penicillin G (13)	15	22	83	–	–	–
Phenylacetic acid methyl ester (14)	100	100	89	41	100	100
NIPAB (15)	64	33	100	–	–	–
NIPGB (16)	27	–	46	–	–	–
APNA (17)	9	11	2	–	–	–

a. Numbers in brackets refer to Fig. 1B.
b. Measurements were carried out with cell-free extracts of the respective clones, supplying substrates at a final concentration of 10 mM. Relative activities were calculated with respect to the compound that was converted the fastest by each extract (1.1 mU, pM1; 0.9 mU, pS1; 31.2 mU, pS2; 47.1 mU, pG1; 0.5 mU, pL1; 0.8 mU, pL2) with 1 mU defined as the conversion of 1 nmol substrate min^{-1} mg $protein^{-1}$. –, no conversion.

observed enzyme activity could easily be identified due to the found homologies, whereas pL1 and pL2 did not contain obvious candidates among their potential protein-coding regions. For the verification of the enzyme-coding sequences by subcloning, we therefore focused on pM1, pS1, pS2 and pG1. On pM1, a putative hydrolase homologue was found, preceded by the vector-encoded P_{lac} promoter and a good ribosome binding site, which made the respective ORF the most likely amidase-coding region although the resulting enzyme would be N-terminally truncated compared with its database relative. The activity of *E. coli* (pS1) on D-phenyglycine-L-leucine, a molecule that is structurally very similar to a dipeptide, indicated the expression of the dipeptidase homologue encoded by ORF 1. Similarly, the substrate profiles of *E. coli* (pS2) and *E. coli* (pG1) supported the assumption that their assayed activities were caused by the encoded penicillin acylase and mandelamide hydrolase homologues respectively (Table 3). Subclones carrying the above-mentioned ORFs in pZero-2 displayed the same activities as the original recombinants, identifying the selected sequences indeed as the enzyme-coding regions. Analysis of the deduced protein sequences suggested that the dipeptidase and the penicillin acylase homologue carry N-terminal signal sequences that may lead to their secretion into the periplasmatic or extracellular space (Table 2).

Substrate profiling

To explore the biocatalytic potential of the retrieved enzymes, their relative activities were determined on a number of different substrates using cell-free extracts (Table 3). All extracts showed no or only low activity towards compounds with an aliphatic acyl moiety, which is not surprising regarding the aromatic substrates used for selection of active clones. Enzymes encoded by plasmid pS1 and pL1 appeared to not accept substrates with a substituent on the Cα-position of the acyl moiety, although *E. coli* (pS1) could grow on D-phenylglycine-L-leucine. Also the pM1 encoded enzyme preferred substrates with no substituent at this position. Recombinants carrying pG1, pS2 and pL2, in contrast, could readily convert α-amino compounds with pG1 and pS2 even being able to accept a α-hydroxy group. Whereas *E. coli* (pG1) was most active on D-phenylglycine amide, all other enzymes hydrolysed the chemically more labile ester bond of phenylacetic acid methyl ester the fastest.

Concerning the nature of the leaving groups of their substrates, enzymes fell into two categories, one being solely active on primary amides or amides carrying a small aliphatic leaving group (pG1, pL1 and pL2), and another group also hydrolysing substrates with more bulky leaving groups, i.e. penicillin G and the colorimetric substrates (pM1, pS1, and pS2). Of the latter group, only *E. coli* (pS2) was found to be capable of both hydrolysing penicillin G and catalysing a condensation reaction when incubated with 6-aminopenicillanic acid and phenylacetamide, yielding penicillin G (Fig. 4). Two side reactions, however, were also found to take place when synthesizing penicillin G, namely the hydrolysis of phenylacetamide and the hydrolysis of the formed antibiotic, which is typical for penicillin amidase-catalysed antibiotic synthesis reactions with activated side-chain donors (Kasche, 1986). As a consequence, the concentration of antibiotic passes through a maximum in time, of which the height is solely determined by the substrate concentrations and the kinetic properties of the enzyme (Youshko and Svedas, 2000). Compared with the well-studied *E. coli* penicillin amidase, the pS2-encoded enzyme allowed a more than 2-fold higher maximal level of penicillin G accumulation

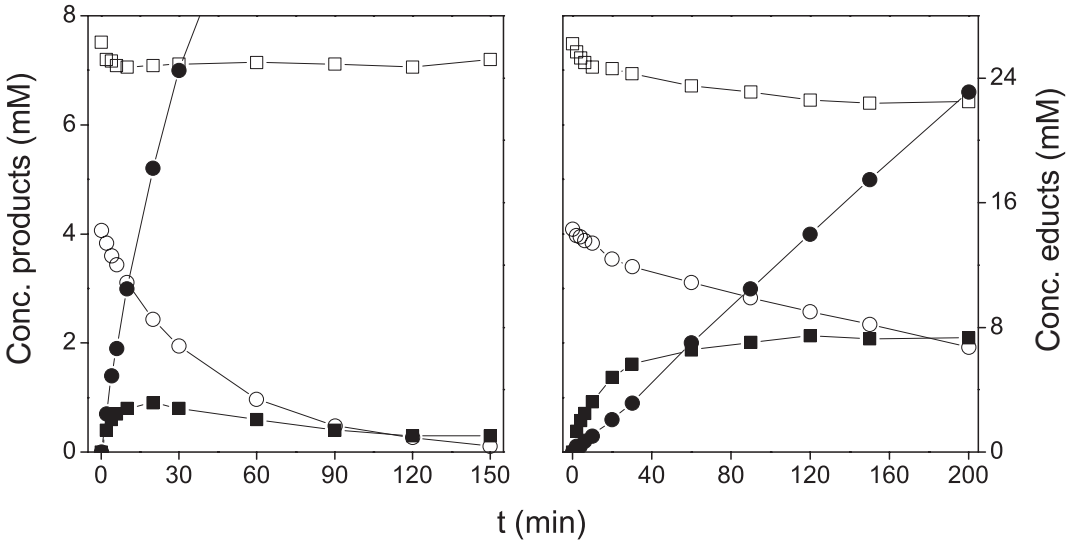

Fig. 4. Biocatalytic synthesis of penicillin G by the penicillin amidase encoded on pS2 (right) and the one of *E. coli* ATCC 11105, carried by the plasmid pEC (left). Cell-free extracts of *E. coli* TOP10 (pS2) and *E. coli* HB101 (pEC) were prepared as described in *Experimental procedures*, except that *E. coli* HB101 cells were grown at 17°C for 3 days with 0.1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) added for induction. Extracts were incubated with 15 mM phenylacetamide and 25 mM 6-aminopenicillanic acid at 30°C. Educt and product concentrations in the course of the reaction were determined by HPLC analysis. The pEC plasmid was kindly provided by DSM (Delft, the Netherlands). Symbols: phenylacetamide (○), 6-aminopenicillanic acid (□), phenylacetic acid (●), penicillin G (■).

(Fig. 4). The performance of the new biocatalyst was also significantly better in the production of the clinically relevant antibiotics ampicillin and amoxicillin, which makes the recovered enzyme a potential candidate for biotechnological application.

In conclusion, cloning from enrichment cultures was found to be an efficient tool for enzyme discovery although cultivation steps preceding DNA extraction may limit the access to certain parts of the microbial gene pool. Taking into account that virtually all experimental steps from DNA extraction to construction and transformation of recombinant plasmids can be performed more efficiently and with less effort, the preparation of several small enrichment gene banks of limited diversity may be preferable compared with the construction of a single exhaustive gene bank without cultivation, especially when time and resources are limited.

Experimental procedures

Environmental samples

Inocula for enrichment cultures were obtained from four different locations in the Netherlands. Marine sediment was collected from the surface of daily inundated saline mud flats during low tide near Paesens-Moddergat. Soil samples were taken from the upper 5–15 cm of the shore of a small goose pond, the lakeshore of the Lauwersmeer (sandy soil), and from a local agricultural field (loamy soil). The latter soil served also for DNA extraction without preceding cultivation.

Media and cultivation conditions

All selective media were based on minimal medium of pH 7.0. It contained per litre 5.3 g $Na_2HPO_4{\cdot}12H_2O$, 1.4 g KH_2PO_4, 0.2 g $MgSO_4{\cdot}7H_2O$, 1.0 g $(NH_4)_2SO_4$, 1 ml vitamin solution (Janssen *et al.*, 1984), and 5 ml of a trace element solution. The trace element solution consisted of 780 mg $Ca(NO_3)_2{\cdot}4H_2O$, 200 mg $FeSO_4{\cdot}7H_2O$, 10 mg $ZnSO_4{\cdot}7H_2O$, 10 mg H_3BO_3, 10 mg of $CoCl_2{\cdot}6H_2O$, 10 mg $CuSO_4{\cdot}5H_2O$, 4 mg $MnSO_4{\cdot}H_2O$, 3 mg $Na_2MoO_4{\cdot}2H_2O$, 2 mg $NiCl_2{\cdot}6H_2O$, and 2 mg $Na_2WO_4{\cdot}2H_2O$ per litre. In nitrogen-free minimal medium $(NH_4)_2SO_4$ was replaced by Na_2SO_4, and $Ca(NO_3)_2{\cdot}4H_2O$ was omitted from the trace element solution. Solutions of carbon, nitrogen, and leucine sources were prepared in 50 mM potassium phosphate buffer (pH 7.0). All media used for gene bank screening and cultivation of isolated clones were supplied with 10 mg l^{-1} kanamycin. When solid LB-containing media were prepared, 15 g l^{-1} agar (Difco) was added. For solid media based on minimal medium, 15 g l^{-1} MP agarose (Roche) was used. Cells were grown at 30°C with shaking of liquid cultures at 200 r.p.m.

Enrichment cultures

Enrichments were carried out under nitrogen-limiting conditions, supplying one or several aromatic and aliphatic amides as sole source of nitrogen. Inocula were prepared by suspending 5 g of environmental material in 50 ml enrichment medium (nitrogen-free minimal medium supplied with 10 mM glucose, 2.5 mM glycerol, and 2.5 mM sodium acetate as carbon sources). The mixtures were homogenized in a standard blender (Moulinex) for three 1 min intervals with 1 min breaks in-between to allow cooling, and incubated overnight

at 30°C and 200 r.p.m. for depletion of residual nitrogen. After settling of coarse material, subsamples of 200 µl were used to inoculate 20 ml of enrichment medium, supplied with 1 mM D-phenylglycine amide as sole source of nitrogen for enrichment cultures ENR-M, ENR-S, and ENR-L. To ENR-G, an equimolar mixture of 10 different amides (Fig. 1B) was added, providing each compound at a final concentration of 0.1 mM. After 2 days of incubation, 0.5 ml of cells were transferred to 50 ml of the same but fresh medium, and incubated for another 2 days before they were pelleted and used for DNA extraction.

DNA extraction

The protocol supplied with the Wizard Genomic DNA Purification Kit for DNA isolation from pure bacterial cultures (Promega) was adapted for the use of enrichment cultures. Cell pellets from enrichment cultures were resuspended in a total volume of 2.5 ml lysis buffer (50 mM Tris-HCl, 50 mM EDTA, 2 mg ml^{-1} lysozyme, and 0.1 mg ml^{-1} proteinase K; pH 7.0) and incubated at 37°C for 30 min. To the suspensions, 3 ml of Nuclei Lysis Solution (Promega) was added. After incubation at 80°C for 15 min, the mixtures were incubated at 65°C for 1–2 h until complete lysis was reached. RNA was subsequently removed by adding 10 µl of DNAse-free RNAse solution (10 mg ml^{-1}) and incubating at 37°C for 30 min. Proteins were precipitated, using 1 ml of Protein Precipitation Solution (Promega), vortexing for 20 s at maximum speed, and incubation on ice for 15 min, followed by a 10 min centrifugation step at 3300 *g*. The supernatants were transferred to 0.1 volumes of sodium acetate (3 M, pH 5.0) and 2 volumes of absolute ethanol. Threads of precipitated genomic DNA were collected with a hooked Pasteur pipette and washed with 70% ethanol. Nucleic acids were resuspended in 0.5–1 ml of 10 mM Tris buffer (pH 8.0) and used directly for gene bank construction.

For DNA isolation directly from soil, a cell extraction-based protocol, employing blending and differential centrifugation to remove cells from the sample matrix prior to lysis, was used as described previously (Gabor *et al*., 2003). DNA prepared this way needed to be further purified on 3% (w/v) agarose gel (Roche), using the QIAEX II Gel Extraction System (Qiagen).

PCR-DGGE analysis

PCR-DGGE analysis was performed as described before (Gabor *et al*., 2003). Primers U968 (5′- AAC GCG AAG AAC CTT AC-3′) and L1401 (5′-CGG TGT GTA CAA GAC CC-3′) served to amplify a 402 bp section of bacterial 16S rRNA genes, including the highly variable V6 region (Engelen *et al*., 1995). Separation of PCR products by denaturing gradient gel electrophoresis (DGGE) was carried out with a device manufactured by Ingeny International BV (the Netherlands). Gels were silver stained (Bassam *et al*., 1991) and digitally photographed.

Gene bank construction

DNA fragments of 4–6 kb were prepared by mechanical shearing of DNA with a nebulizer (Invitrogen, catalog no. 7025–05) according to the instructions of the manufacturer. After ethanol precipitation in presence of 50 µg ml^{-1} glycogen as a carrier, DNA was resuspended and blunted with Klenow and T4 polymerase (Invitrogen). Fragments were purified from agarose gel using the QIAquick Gel Extraction Kit (Qiagen), and ligated into the *Eco*RV site of the high-copy plasmid vector pZero-2 (Invitrogen), using a 1 : 10 vector:insert ratio and a total amount of about 1 µg of DNA in a volume of 50 µl. T4 ligase (Invitrogen) was added to a final concentration of 0.1 U μl^{-1} and ligation mixtures were incubated at 16°C for 1 h before ethanol precipitation. Precipitated DNA was washed with 70% ethanol and resuspended in 10 µl demineralized water. Aliquots of 2.5–5 µl were transformed to 80 µl of *E. coli* TOP10 cells [Δ(ara-leu)7697] by electroporation, using a 2 mm gap electroporation cuvette (Invitrogen) and a Gene Pulser apparatus (Bio-Rad). Electrocompetent cells were prepared according to standard procedures (Sambrook *et al*., 1989) and allowed the production of $>10^9$ clones μg^{-1} supercoiled control plasmid (pSK^-) and 8000 (loamy soil DNA) to 25 000 clones (enrichment DNA) when transforming ligation mixtures as described above.

Transformants were spread on LB agar plates containing 10 mg l^{-1} kanamycin for the selection of transformants carrying a pZero-2 derived recombinant plasmid. After incubation at 37°C for 1 day, colony forming units (cfu) were enumerated and at least 20 transformants were separately grown overnight in 5 ml LB medium. Plasmid DNA was isolated from these cultures using the High Pure Plasmid Isolation Kit (Roche) and checked for insert size by enzymatic restriction with *Bam*HI and *Xho*I (Roche). Average insert sizes were found to be 5.2 kb in all gene banks, with background levels of self-ligated vector molecules of ≤5%.

Selection of amidase expressing clones

Transformants were washed from plates with minimal medium. Aliquots of the amplified gene banks were spread on minimal medium plates supplied with 0.2% (w/v) glucose and 10 mg l^{-1} phenylacetyl-L-leucine or D-phenylglycine-L-leucine as the sole source of leucine. Colonies of amidase-expressing clones were visible after 3 days of incubation at 30°C. Aliquots of the gene bank derived from ENR-G were plated on 10 different nitrogen-free minimal media, containing 0.2% (w/v) glucose, 5 mg l^{-1} leucine, and one of the amides used for enrichment (Fig. 1) as the only nitrogen source (250 µM final concentration). Positives could be scored after 3–5 days of growth at 30°C. Because of the use of amplified gene banks, active clones obtained after selection needed to be analysed by enzymatic restriction to determine the number of unique transformants exhibiting the respective enzymatic activity.

HPLC analysis

All high-performance liquid chromatography (HPLC) analyses were carried out using a 10 cm long Chrompack C18 column (0.5 cm diameter) in connection with Jasco PU-980 pumps and a Jasco MD-910 detector set at 214 nm. Compounds were isocratically eluted at a flow rate of 1 ml min^{-1} with a solution of 340 mg l^{-1} sodium dodecylsulphate and

680 mg l^{-1} KH_2PO_4.3 H_2O in a 30 : 70 (v/v) acetonitrile/water mixture of pH 3.0 (adjusted with diluted phosphoric acid). Before injection, samples were quenched by 1 : 50 dilution in eluent.

Substrate profiling

All substrate conversions were carried out with cell-free extracts that were prepared from overnight cultures grown at 30°C and 200 r.p.m. rotary shaking. Cells were harvested by centrifugation at 5000 *g* for 10 min. Pellets were resuspended in 1/20 volume of ice-cold 50 mM potassium phosphate buffer (pH 7.0) containing 0.1 mM 1,4-dithiothreitol, and disrupted by sonification. Supernatants obtained after centrifugation at 5000 *g* for 30 min at 4°C were used for the measurements. Protein concentrations of the cell-free extracts were determined by the method of Bradford. Enzymatic hydrolysis of the colorimetric substrates (Fig. 1B, compounds 15–17) was followed by measuring the increase in absorbance at 405 nm caused by the release of 5-amino-2-nitrobenzoic acid ($\Delta\varepsilon_{405nm}$ = 9.09 mM^{-1} cm^{-1}) or *p*-nitroaniline ($\Delta\varepsilon_{405nm}$ = 13 mM^{-1} cm^{-1}), respectively, using a Perkin Elmer Lambda Bio 40 spectrometer. Conversion of non-chromogenic substrates was monitored by withdrawing samples of the reaction mixtures at different time points and analysing them by HPLC. All substrates were supplied at a final concentration of 10 mM and reactions were carried out at 30°C in 50 mM potassium phosphate buffer of pH 7.0. Background activities caused by host strain proteins were determined with cell-free extract of *E. coli* TOP10 and subtracted from the activities measured with the recombinant *E. coli* strains.

Coding sequence determination and analysis

DNA insert sequences were determined at the Department of Medical Biology of the University of Groningen (the Netherlands) or BaseClear Holding B.V. (Leiden, the Netherlands) with at least two times coverage of each base. Open reading frames (ORFs) were identified with the ORF Finder service available at the NCBI web page (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). For translation to protein sequences, the Bacterial Code was selected, allowing ATG, GTG, TTG, ATT and CTG as alternative start codons. The minimal ORF length was set to 270 bp. When alternative start sites for the same gene were encountered, the largest ORF was selected. ORFs were translated and used as queries in BLAST searches of the DDBJ/EMBL/GenBank database entries. Translated ORFs were analysed for the presence of N-terminal signal sequences, using the SignalP program (Henrik *et al*., 1997).

The sequence data have been submitted to the DDBJ/EMBL/GenBank databases under accession numbers AY573296 (plasmid pM1), AY573297 (plasmid pS1), AY573298 (plasmid pS2), AY573299 (plasmid pG1), AY573300 (plasmid pL1) and AY573301 (plasmid pL2).

Acknowledgements

This work was financed by STW under project GCH.4919. We thank M. van der Velde and S. den Dulk from Bioclear (Groningen, the Netherlands) for DGGE analysis.

References

Bassam, J.B., Caetano-Anollés, G., and Gresshof, P.M. (1991) Fast and sensitive silver staining of DNA in polyacrylamide gels. *Anal Biochem* **196:** 80–83.

Bodenteich, A., Chissoe, S., Wang, Y.-F., and Roe, B.A. (1994) *Automated DNA Sequencing and Analysis*. Venter, C. (ed.). San Diego, USA: Academic Press.

Bruggink, A., Roos, E.C., and de Vroom, E. (1998) Penicillin acylase in the industrial production of β-lactam antibiotics. *Org Proc Res Dev* **2:** 128–133.

Cottrell, M.T., Moore, J.A., and Kirchman, D.L. (1999) Chitinases from uncultured marine microorganisms. *Appl Environ Microbiol* **65:** 2553–2557.

Courtois, S., Cappellano, C.M., Ball, M., Francou, F.-X., Normand, P., Helynck, G., *et al.* (2003) Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. *Appl Environ Microbiol* **69:** 49–55.

Daniel, R. (2002) Construction of environmental libraries for functional screening of enzyme activity. In *Directed Molecular Evolution of Proteins*. Brakmann, S., and Johnsson, K., (eds). Weinheim, Germany: Wiley-VCH, pp. 63–77.

Engelen, B., Heuer, H., Felske, A., Nübel, U., Smalla, K., and Backhaus, H. (1995) Protocols for TGGE. In *Abstracts for the Workshop on Application of DGGE and TGGE in Microbial Ecology*. Braunschweig, Germany: Biologische Bundesanstalt für Land- und Forstwirtschaft.

Entcheva, E., Liebl, W., Johann, A., Hartsch, T., and Streit, W.R. (2000) Direct cloning from enrichment cultures, a reliable strategy for isolation of complete operons and genes from microbial consortia. *Appl Environ Microbiol* **67:** 89–99.

Forney, L.J., and Wong, D.C.L. (1989) Alteration of the catalytic efficiency of penicillin amidase from *Escherichia coli*. *Appl Environ Microbiol* **55:** 2556–2560.

Gabor, E.M., de Vries, E.J., and Janssen, D.B. (2003) Efficient recovery of environmental DNA for expression cloning by indirect extraction methods. *FEMS Microbiol Ecol* **44:** 153–163.

Gelsomino, A., Keijzer-Wolters, A.C., Cacco, G., and van Elsas, J.D. (1999) Assessment of bacterial community structure in soil by polymerase chain reaction and denaturing gradient gel electrophoresis. *J Microbiol Meth* **38:** 1–15.

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.

Healy, F.G., Ray, R.M., Aldrich, H.C., Wilkie, A.C., Ingram, L.O., and Shanmugam, K.T. (1995) Direct isolation of functional genes encoding cellulases from the microbial consortia in a thermophilic, anaerobic digester maintained on lignocellulose. *Appl Microbiol Biotechnol* **43:** 667–674.

Henne, A., Schmitz, R.A., Bömeke, M., Gottschalk, G., and Daniel, R. (2000) Screening of environmental DNA libraries for the presence of genes conferring lipolytic activity on *Escherichia coli*. *Appl Environ Microbiol* **66:** 3113–3116.

Henrik, N., Engelbrecht, J., Brunak, S., and von Heijne, G. (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng* **10:** 1–6.

Janssen, D.B., Scheper, A., and Witholt, B. (1984) Biodegradation of 2-chloroethanol by pure bacterial cultures. *Prog Ind Microbiol* **29:** 169–178.

Kasche, V. (1986) Mechanism and yields in enzyme catalyzed equilibrium and kinetically controlled synthesis of β-lactam antibiotics, peptides and other condensation products. *Enzyme Microb Technol* **8:** 4–16.

Knietsch, A., Waschkowitz, T., Bowien, S., Henne, A., and Daniel, R. (2003) Construction and screening of metagenomic libraries derived from enrichment cultures: generation of a gene bank for genes conferring alcohol oxidoreductase activity on *Escherichia coli*. *Appl Environ Microbiol* **69:** 1408–1416.

Lorenz, P., and Schleper, C. (2002) Metagenome – a challenging source of enzyme discovery. *J Mol Catal B-Enzym* **19–20:** 13–19.

Lorenz, P., Liebeton, K., Niehaus, F., and Eck, J. (2002) Screening for novel enzymes for biocatalytic processes: accessing the metagenome as a resource of novel functional sequence space. *Curr Opin Biotechnol* **13:** 572–577.

Oefner, P.J., Hunicke-Smith, S.P., Chiang, L., Dietrich, F., Mulligan, J., and Davis, R.W. (1996) Efficient random subcloning of DNA sheared in a recirculating point-sink flow system. *Nucleic Acids Res* **24:** 3879–3886.

Rondon, M.R., August, P.R., Bettermann, A.D., Brady, S.F., Grossman, T.H., Liles, M.R. *et al.* (2000) Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl Environ Microbiol* **66:** 2541–2547.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor, USA: Cold Spring Harbor Laboratory Press.

Santegoeds, C.M., Nold, S.C., and Ward, D.M. (1996) Denaturing gradient gel electrophoresis used to monitor the enrichment culture of aerobic chemoorganotrophic bacteria from a hot spring cyanobacterial mat. *Appl Environ Microbiol* **62:** 3922–3928.

Schink, B. (2002) Synergistic interactions in the microbial world. *Antonie Van Leeuwenhoek* **81:** 257–261.

Tebbe, C.C., and Vahjen, W. (1993) Interference of humic acids and DNA extracted directly from soil in detection and transformation of recombinant DNA from bacteria and yeast. *Appl Environ Microbiol* **59:** 2657–2665.

Torsvik, V., Øvreås, L., and Thingstad, T.F. (2002) Prokaryotic diversity – magnitude dynamics, and controlling factors. *Science* **296:** 1064–1066.

Youshko, M.I., and Svedas, V.K. (2000) Kinetics of ampicillin synthesis catalyzed by penicillin acylase from *E. coli* in homogeneous and heterogeneous systems. Quantitative characterization of nucleophile reactivity and mathematical modeling of the process. *Biochemistry* **65:** 1367–1375.