Available online at www.sciencedirect.com
SCIENCE DIRECT®
Biochemical and Biophysical Research Communications 335 (2005) 57–65
BBRC
www.elsevier.com/locate/ybbrc

Isolation of novel lipolytic genes from uncultured bacteria of pond water

Ravi Ranjan [a], Amit Grover [a], Raj Kishor Kapardar [b], Rakesh Sharma [a,b,*]

[a] Department of Environmental Biology, University of Delhi, Delhi 110007, India
[b] Institute of Genomics and Integrative Biology, Mall Road, Delhi 110007, India

Received 1 July 2005
Available online 20 July 2005

Abstract

Metagenomic libraries give access to gene pool of bacteria present in environmental samples avoiding the culture bias. A metagenomic library of pond water microbial assemblage in plasmid vector containing about 532 Mb of community DNA was prepared. Screening of a part of the unamplified library resulted in isolation of 11 unique lipolytic clones with an ability to hydrolyze tributyrin. DNA sequence of the lipolytic genes varied in G + C composition from 57% to 75%. Twelve lipolytic genes encoding proteins with 25–70% amino acid identity with proteins in the databases were identified. Ten of the encoded proteins belonged to seven known lipolytic protein families. One of the proteins was similar to recently identified esterase BioH. A lipolytic protein with high similarity to yet uncharacterized α/β hydrolase protein family abh_upf0017 was identified from one of the clones. Conserved motif for lipolytic enzymes GXSXG, conserved aspartic and histidine residues were identified in this encoded protein.

Keywords: Metagenomic; Metagenome; Lipolytic; Lipase; Esterase; α/β-hydrolase; Uncultured bacteria; Unculturable; Culture-independent; Pond water

Lipolytic enzymes include lipases (EC 3.1.1.3), which hydrolyze long chain acylglycerols ($\geqslant C_{10}$) and esterases (EC 3.1.1.1), which hydrolyze short chain acylglycerol ($\leqslant C_{10}$). Lipases are also able to hydrolyze short chain acylglycerols [1]. Lipolytic enzymes have significant biotechnological importance because of their ability to catalyze regio- and stereo-selective organic reactions [1,2]. These enzymes are also becoming important in synthetic organic chemistry because of their capability to withstand organic solvents and catalyze reverse reactions [2]. Identification of lipolytic enzymes with a spectrum of sequence and functional diversity will enrich the toolbox of synthetic chemists, which will help in quick selection of a suitable biocatalyst for challenging reactions and reaction conditions [3,4]. Large number of lipases and esterases has been identified from different domains of life. Lipase engineering database contains more than 800 protein entries assigned to 38 homologous families and 16 superfamilies [5]. Bacterial lipolytic enzymes have been classified in eight families by Arpigny and Jaeger [6]. Availability of large number of complete genome sequences is augmenting the discovery of newer lipolytic enzymes and lipolytic gene families [7–9].

Majority of the bacteria present in environment cannot be cultured in laboratory by traditional culture techniques [10]. This vast diversity of 'yet to be cultured' bacteria represents a large gene pool for biotechnological exploitation. Culture-independent methods like preparation of metagenomic libraries have allowed successful exploitation of microbial communities from various environmental niches [11–13]. Screening of sufficiently large metagenomic libraries can result in isolation of many genes of similar function with significant sequence variation [14,15]. Functional lipolytic genes have been previously identified from metagenomic libraries of soil [11,16,17], hot spring sediments [13], and alkaline soda lakes [18]. PCR based strategy was

* Corresponding author. Fax: +91 11 27667471.
E-mail address: rsharma@igib.res.in (R. Sharma).

0006-291X/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.bbrc.2005.07.046

used to identify a novel lipolytic gene from soil metagenome [19], which avoids the reported difficulties in functional expression of lipolytic genes in heterologous host [2].

Fresh water ponds, lakes, and water bodies have been shown to possess unique microbial diversity including members of various unculturable groups [20,21]. To the best of our knowledge, fresh water microbial communities have not been exploited by culture-independent methods for isolation of lipolytic genes. Here, we report construction of a metagenomic library from pond water microbial assemblage and identification of 11 unique lipolytic clones with an ability to hydrolyze tributyrin.

Materials and methods

Sample, strains, and culture conditions. Pond water sample was collected from University of Delhi campus in month of June 2002. It was greenish-brown in colour and had a pH of 8.5. Sample was collected in a sterile container and cells were harvested immediately for isolation of metagenomic DNA. *Escherichia coli* DH10B was a gift from Dr. H. Shizuya, Caltech. The *E. coli* strains were grown in Luria–Bertani (LB) liquid medium or on LB agar plates at 37 °C. The media were supplemented with 50 μg kanamycin and/or 100 μg ampicillin per ml for selection of plasmids.

Pond water metagenomic DNA isolation. Microbial cells from 5 L of pond water sample were harvested by centrifugation at 8000*g* for 10 min. Metagenomic DNA was isolated with modification of the method described by Zhou et al. [22]. The cell pellet (greenish in colour) was suspended in 20 ml of extraction buffer {100 mM Tris–HCl [pH 8.0], 100 mM EDTA, 100 mM sodium phosphate, 1.5 M NaCl, 1% cetyltrimethylammonium bromide (CTAB)} and incubated at 37 °C for 15 min. Lysozyme was added at a final concentration of 5 mg/ml and tubes were incubated at 37 °C for 1 h with gentle shaking. Proteinase K (2.5 mg) and SDS at a final concentration of 1.0% were added, and the tubes were incubated at 65 °C for 1 h. The lysate was centrifuged at 8000*g* for 10 min and the supernatant was transferred to a fresh tube. It was then extracted once with equal volume of 50:50 phenol and chloroform. The DNA was precipitated from aqueous phase by 0.6 V of isopropyl alcohol and kept undisturbed for 1 h. The DNA was obtained by centrifugation at 10,000*g* for 20 min and washed with 20 ml of 70% ethanol. The pellet was air dried and dissolved in 15 ml of 10 mM Tris and 1 mM EDTA buffer (TE), pH 8.0. DNA was further purified by CTAB precipitation [23] and dissolved in 3.0 ml TE, pH 8.0.

Metagenomic library construction. Ten-microgram pond metagenomic DNA was partially digested with *Sau*3A1 and DNA fragments in the range of 2–12 kb were gel eluted using gel band elution kit (Amersham Biosciences, UK). Two hundred nanogram of partially digested DNA was ligated to *Bam*HI digested and dephosphorylated pUC19 vector. The ligated product was electroporated at 200 Ω, 25 μf, and 2.5 kV in *E. coli* DH10B using Micropulser II (Bio-Rad, USA). Transformants were obtained on ampicillin containing LB agar plates.

Screening of lipolytic clones. Lipolytic clones were screened by plating transformation product on LB ampicillin plates containing 1.0% tributyrin [24]. The lipolytic clones were selected based on their ability to produce clear halo around the colony after 48 h incubation at 37 °C. All lipolytic clones obtained after preliminary screening were streaked to obtain single colonies and confirmed again for production of hydrolysis zone on tributyrin containing plates.

DNA manipulation techniques. Standard procedure of plasmid isolation, restriction enzyme digestion, ligation, and competent cell preparation were used as described by Sambrook and Russell [25] and Ausubel et al. [23].

Transposon mutagenesis and mapping. Plasmid from the unique clones was used for in vitro transposon mutagenesis by Template Generation System (TGS, F-700) kit by the manufacturer's provided protocol (Finnzyme, Finland). The mutagenized plasmid was electroporated in *E. coli* DH10B and colonies containing transposon-mutated plasmids were selected on ampicillin and kanamycin containing LB agar plates. These colonies were then stabbed with the help of sterile toothpicks on 1.0% tributyrin, ampicillin, and kanamycin containing LB agar plates to select negative mutants for lipolytic activity. Fifty to 200 mutants were screened for each clone to obtain negative mutants. These mutants were expected to contain transposon insertion in lipolytic gene or in their regulatory region. Location of the transposon insertion in the selected mutated plasmids was mapped by size of PCR product obtained by primers against transposon ends and pUC reverse or pUC forward primer. Mutants with insertions separated by a distance of about 400–450 bp were selected for sequencing. Assembly of these sequences in most cases resulted in double strand sequence of the lipolytic gene region and single strand sequence from the ends and flanking region of lipolytic genes. Primers at the extreme ends of the single strand sequences were used to obtain double strand sequence for both the ends.

DNA sequencing and analysis. DNA sequencing reactions were done with big dye termination cycle sequencing kit version 3.1 (Applied Biosystems, USA) and the products were run on an ABI 3700 machine using POP6 method at The Centre for Genomic Application (TCGA), New Delhi, India. Transposon end specific primers SeqA and SeqB (supplied with the TGS kit) were used to sequence mutated plasmids. Sequence assembly and analysis were carried out using Lasergene package, version 5.07 (DNA Star, USA). Open reading frames in the assembled sequence for each clone were identified by ORF finder [26] at National Centre for Biotechnology Information (NCBI) website and the amino acid sequence of each identified ORF was used to find the closest match by BLAST without low complexity filter [27]. Encoded proteins were analyzed by BLAST for similarity with known lipolytic proteins and α/β hydrolase fold containing proteins at Lipase Engineering Database [5] and Esther database [28], respectively. Multiple sequence alignments were carried out by ClustalW [32]. Conserved domains and patterns were analyzed using conserved protein domain database [29], InterProScan [30], and Prosite [31]. Signal peptide and transmembrane domain were predicted using server SignalP [33] and HMMTOP [34], respectively.

Nucleotide sequence accession number. The sequences obtained from different clones were deposited in GenBank under Accession Nos. DQ077738–DQ077748.

Results

Construction of pond water metagenomic library

Water from a small pond in University of Delhi campus was collected. The surface of the water was covered with algal mass, which was avoided during collection of the water sample. The water appeared greenish-brown in colour and had a pH of 8.5. Microscopic observation revealed presence of bacterial, cyanobacterial, filamentous algal cells, and few possible members of invertebrate groups. Microbial cells from 5 L of pond water were harvested immediately after the collection by centrifugation. The metagenomic DNA obtained after initial isolation procedure was resistant to digestion with restriction enzymes. It was further purified by CTAB to obtain readily digestible pure DNA. CTAB is known to remove polysaccharides and humic acid impurities [22]. Pond

water metagenomic library was constructed containing more than 1,40,000 recombinant colonies. More than 90% of the colonies in the library were recombinant. Analysis of insert fragments generated by *Eco*RI and *Hin*dIII restriction digestion of 40 recombinant plasmids was used to estimate an average insert size of ~3.8 kb for the library (data not shown). The metagenomic library represented about 532 Mb of the pond water microbial community DNA.

Screening of lipolytic clones

A part of unamplified library (~30,000 colonies) was screened for lipolytic clones on 1.0% tributyrin containing LB ampicillin plates. Seventeen clones were obtained in the initial screening for their ability to produce hydrolysis zone after 48 h of incubation at 37 °C. Only 13 clones gave stable activity after single colony re-streaking. The zone size of hydrolysis of tributyrin by different clones varied from 4 to 15 mm after 72 h incubation at 37 °C, indicating variable expression or substrate preference of the lipolytic enzymes produced by the clones (Fig. 1). As no IPTG was added to the screening plates, the activity obtained was expected to be because of the expression from the native promoter in the clones or from the low level leaky expression from *lac* promoter. All the colonies obtained after retransformation of the plasmids isolated from the stable clones produced lipolytic activity indicating plasmid-borne nature of the lipolytic activity. Plasmids from the stable clones were designated as pLR1–pLR13. The insert size in these plasmids varied from 2.2 to 8.5 kb (Fig. 2). Restriction digestion analysis of the 13 plasmids revealed that plasmid pLR9 was a sibling of pLR11 and pLR10 of pLR13 (data not shown). Hence, pLR11 and pLR13 were not used for further analysis.

To analyze the substrate specificity of the lipolytic enzymes produced by the clones, the clones were stabbed in to 1.0% tributyrin, or olive oil and rhodamine B containing LB agar plates [35]. Most of the clones produced hydrolysis zone of varying sizes on tributyrin containing plate within 72 h of incubation at 37 °C but they failed to produce any fluorescent halo on olive oil and rhodamine B containing plates even after 7 days of incubation. These results indicated that the lipolytic enzymes produced by the clones were esterases.

Sequence analysis

DNA sequence of the lipolytic gene region was obtained by sequencing the transposon mutants with no lipolytic activity (Fig. 2) and primer walking. The G + C content of the sequenced region from different clones varied from 56% to 75% (data not shown) and nucleotide BLAST of these sequences did not result in any significant match. This indicated the origin of the inserts from varied bacterial phylotypes not closely related to the bacteria for which the sequence information is available in databases. DNA sequence analysis of lipolytic clones resulted in identification of candidate lipolytic gene from each plasmid. In one plasmid, pLR12, two tandem genes with similarity to lipolytic genes were identified. The proteins encoded from these genes showed 25–70% identity at amino acid level with the

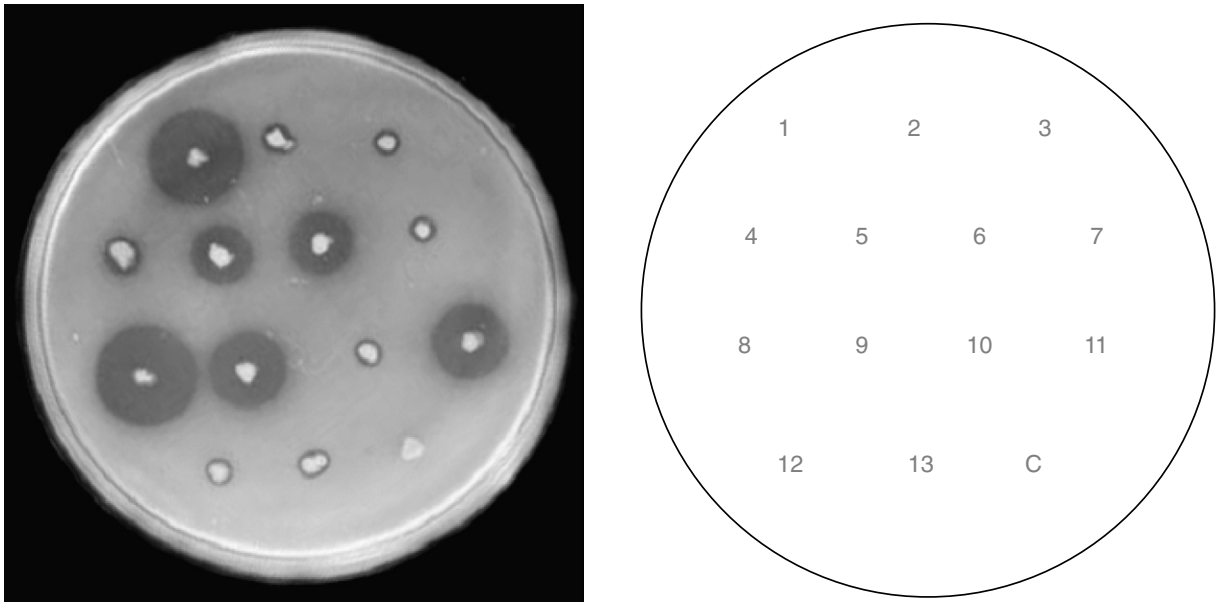

Fig. 1. Lipolytic activity of different clones isolated from pond water metagenomic library on 1.0% tributyrin containing LB agar. Activity was observed after 72 h of incubation at 37 °C. Side panel represents the position of clones on the tributyrin plate, 1–13; pLR1/DH10B to pLR13/DH10B and C; pUC19/DH10B.

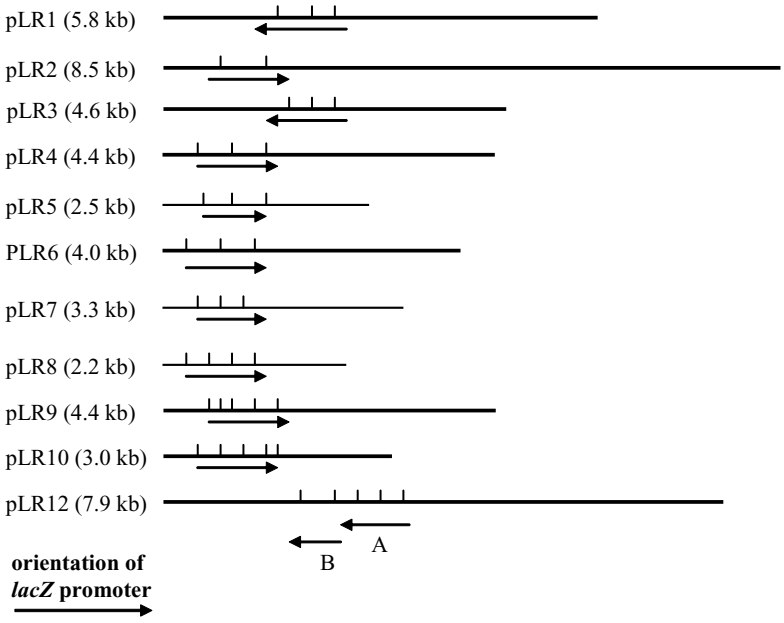

Fig. 2. Schematic representation of the organization of the lipolytic clones pLR1-pLR10, and pLR12 with respect to pUC19 *lacZ* promoter. Insert sizes are mentioned in brackets. Horizontal arrow represents the lipolytic ORF. Vertical lines in the insert represent the site of transposon insertion in the region of lipolytic gene. Note that pLR12 harbors two tandem lipolytic genes A and B.

proteins from *Acidovorax* sp., *Azoarcus* sp., *Bacillus niacini*, *Bradyrhizobium japonicum*, *Brassica oleracea*, *Dechloromonas aromatica*, *Rubrivivax gelatinosus*, *Streptomyces coelicolor*, *Xanthomonas axonopodis*, and uncultured bacterium (Table 1). These results indicated presence of DNA from wide variety of unique bacterial phylotypes in the pond water metagenomic library.

DNA sequence of 2402 bp obtained from the selected region of the pLR1 contained an ORF, which encoded a protein of 385 amino acids. This encoded protein was most similar to a hypothetical protein blr5441 from *B. japonicum* with 52% amino acid identity. It had a complete domain for β-lactamases. Close matches of lipolytic protein from pLR1 contained β-lactamase domain and many of them were annotated as putative β-lactamases or penicillin-binding proteins in the database. We found no significant match for this protein in lipase engineering database but comparison with bacterial lipolytic enzymes revealed that it belonged to β-lactamase fold containing bacterial esterase family VIII of lipolytic enzymes [6]. Multiple sequence alignment of protein from pLR1 with members of family VIII revealed the presence of conserved β-lactamase SXXK motif and another conserved motif LLXHXXG in the encoded protein (Fig. 3A).

An encoded protein of 277 amino acids from pLR2 showed maximum similarity with chlorophyllase 2 from *B. oleracea* with 25% amino acid identity. The other

Table 1
Properties of the lipolytic proteins from pond water metagenome and similar proteins in GenBank

Clone			Homologous protein			
Name	ORF GC (%)	ORF (aa)[a]	Best match (Accession No.)	Organism	*e* value	Identity (%)
LR1	67.36	385	Hypothetical protein (NP_772081)	*B. japonicum*	e−107	204/386 (52)
LR2	72.66	277	Chlorophyllase 2 (AAN51934)	*B. oleracea*	2e−04	38/152 (25)
LR3	71.52	315	Lipase/esterase (AAS77247)	Uncultured bacterium	3e−71	148/312 (47)
LR4	71.07	294	D-(−)-3-Hydroxybutyrate oligomer hydrolase (BAB19271)	*Acidovorax* sp.	e−116	203/290 (70)
LR5	57.71	278	Predicted hydrolases or acyltransferases (YP_158932)	*Azoarcus* sp.	5e−68	131/255 (51)
LR6	75.08	304	Esterase/lipase (ZP_00245626)	*R. gelatinosus*	4e−92	200/286 (69)
LR7	64.05	343	Putative hydrolase (CAC01373)	*S. coelicolor*	7e−45	121/287 (42)
LR8	61.73	458	Esterase 54 (AAQ03995)	*B. niacini*	2e−77	183/473 (38)
LR9	76.15	259	Biotin biosynthesis protein (AAM35277)	*X. axonopodis*	3e−41	114/242 (47)
LR10	61.49	334	Predicted hydrolase (ZP_00348574)	*D. aromatica*	7e−92	177/317 (55)
LR12A	57.73	301	Predicted hydrolases or acyltransferases (ZP_00243877)	*R. gelatinosus*	2e−78	148/295 (50)
LR12B	57.06	287	Predicted esterase (ZP_00152189)	*D. aromatica*	1e−75	144/255 (56)

[a] aa, amino acids.

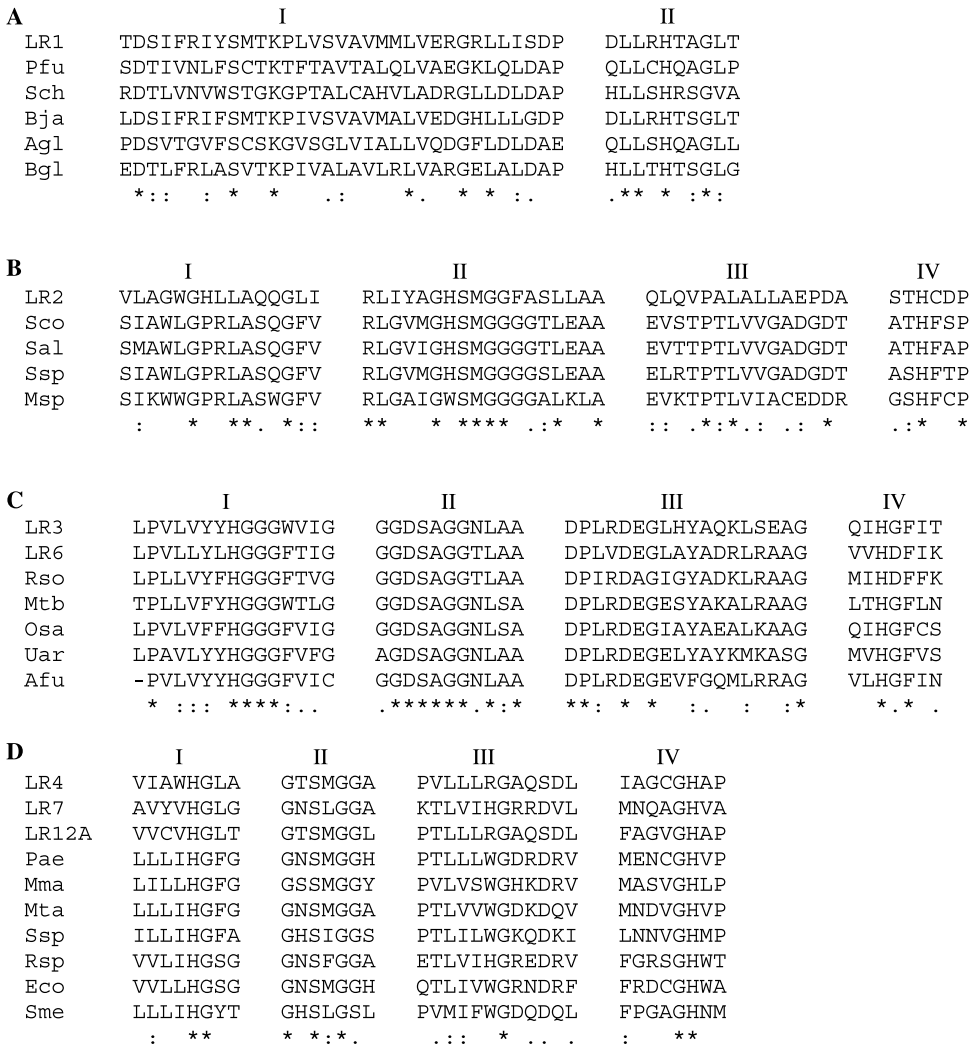

Fig. 3. Conserved sequence blocks from multiple sequence alignment of lipolytic proteins from pond water metagenomic library (A) LR1 # DQ077738 with members of family VIII lipolytic proteins: Pfu, *Pseudomonas fluorescens* SIK W1 # AC60471; Sch, *Streptomyces chrysomallus* # CAA78842; Bja, *B. japonicum* # NP_772081; Agl, *Arthrobacter globiformis* # AAA99492; Bgl, *Burkholderia gladioli* # AAF59826. (B) LR2 # DQ077739 with members of Moraxella lipase 1 family lipolytic proteins: Sco, *S. coelicolor* A3 (2) # CAB69685; Sal, *Streptomyces albus* # AAA53485; Ssp, *Streptomyces* sp. (strain M11) # JN0490; Msp, *Moraxella* sp. TA144 # P19833. (C) LR3 # DQ077740 and LR6 # DQ077743 with members of Moraxella lipase 2 family lipolytic (superfamily hormone sensitive lipase) proteins: Rso, *Ralstonia solanacearum* # CAD15474; Mtb, *Mycobacterium tuberculosis* CDC1551 # AAK47374; Osa, *Oleomonas sagaranensis* # BAA82510; Uar, uncultured archaeon clone estE1 # AY726780; Afu, *Archaeoglobus fulgidus* # 1JJI_D. (D) LR4 # DQ077741, LR7 # DQ077744, and LR12A # DQ077748 with members of Moraxella lipase 3 family lipolytic proteins: Pae, *Pseudomonas aeruginosa* PAO1 # H83276; Mma, *Moritella marina* # AAK07450; Mta, *Moraxella* sp. TA144 # P24640; Ssp, *Streptococcus* sp. N1 # AAK81864; Rsp, *Rhodococcus* sp. # BAA31163; Eco, *E. coli* K-12 # E64762; Sme, *Sinorhizobium meliloti* 1021 # AAK65608. All residues are identical (★), there are conserved substitutions (:) and there are semiconserved substitutions (.) in that column in the alignment.

close matches of this protein were chlorophyllase, predicted dienelactone hydrolase, and lipase from *Arabidopsis thaliana*, *Nostoc punctiforme*, and *S. coelicolor*, respectively. This protein belonged to Moraxella lipase 1 family of lipolytic enzymes (Fig. 3B). Signal peptide and a single transmembrane region from 5th to 22nd amino acid at amino-terminal region of encoded lipolytic protein from pLR2 were predicted indicating that it might be a membrane bound enzyme.

DNA sequence of 2453 bp obtained from the selected region of the pLR3 resulted in identification of an ORF encoding for a protein of 444 amino acids. Homology search revealed best match of this protein with a lipase/esterase from uncultured bacterium of 311 amino acids. Most of the other close matches were also similar in length. Analysis of the sequence for start sites downstream of the predicted start site helped in identification of two more putative start sites at 901 and 1144 bp, which form ORFs encoding for proteins of 396 and 315 amino acids, respectively. A putative ribosome-binding site (RBS) AGGAGA was detected upstream of start codon at 1144 bp. Presence of RBS and matches of similar size proteins in other genomes strongly indicated that the start codon at 1144 bp might be the actual

start site for the lipolytic protein expressed from pLR3. This protein belonged to family Moraxella lipase 2 of hormone sensitive lipases superfamily (Fig. 3C).

Analysis of 1201 bp DNA sequence of pLR4 revealed the presence of an ORF encoding for a protein of 294 amino acid with high similarity to D-(−)3-hydroxy-butyrate oligomer hydrolase of *Acidovorax* sp. with 70% amino acid identity. This protein belonged to family Moraxella lipase 3 of superfamily non-heme peroxidases (Fig. 3D).

A predicted ORF encoding for 198 amino acid long protein from pLR5 showed best match with predicted hydrolase or acyltransferase from *Azoarcus* sp. of 289 amino acids with 51% identity. Search for an alternate start site revealed the presence of a TTG start site 240 bp upstream of the observed start site, coding for a protein of 278 amino acids. A putative ribosome-binding site AAGGAA was observed upstream of this TTG start site indicating that this might be the actual start site of the encoded protein from pLR5. Search in prosite database with encoded lipolytic protein from pLR5 revealed the presence of ribosomal protein S7 signatures [DENSK]-X-[LIVMDET]-X(3)-[LIVMFTA](2)-X(6)-G-K-[KR]-X(5)-[LIVMF]-[LIVMFC]-X(2)-[STAC] at amino-terminal region. This pattern was absent in the close protein matches of lipolytic protein from pLR5 obtained by BLAST in NCBI database. This protein belonged to family non-heme peroxidase of superfamily non-heme peroxidase (Fig. 4A).

A gene encoding 304 amino acid long protein showed best match with putative esterase/lipase of *R. gelatinosus* with 68% identity at amino acid level and was identified from 1437 bp DNA sequence obtained from pLR6. This protein belonged to Moraxella lipase 2 family of hormone sensitive lipase superfamily (Fig. 3C). InterProScan search revealed the presence of HTH_FIS domain at the carboxyl-terminal of this protein. We

```
A              I                II            III        IV       V
LR5     HGAFTAAWTWSEHFL  GWPAHAVSLRGHGGS  GHSMGGFVV  LMCSVPP  GMGHGM
Atu     HGWPLSSDDWDAQML  GFRVVAHDRRGHGRS  GHSTGGGEV  LVSAVPP  GFSHGM
Bpy     HGWPLSGDDWDAQML  GYRVIAHDRRGHGRS  GHSTGGGEV  LVSAVPP  GYSHGM
Rle     HGWPLSSDDWDAQML  GFRVVAHDRRGHGRS  GHSTGGGEV  LVSAVPP  GYPHGM
Ppu     HGWPLSADDWDAQML  GFRVFAHDRRGHGRS  GHSTGGGEV  LIAAVPP  GFPHGM
Pae     HGWPLSSDDWDAQML  GFRVVAHDRRGHGRS  GHSTGGGEV  IISAVPP  GFPHGM
Rer     HGWPLSSDDWDSQLL  GYRVIAHDRRGHGRS  GHSTGGGEV  LIGAVPP  GLSHGM
        **   :.  *. ::*  *: . * . **** *  *** **  *  :: :***  *  ***

B         I       II                 III              IV       V
LR8     RFEAP  HGGAYLNG  VTIFGESAGGSAVLSLMAAPSATGLF  FAAKG  RSTLRID
Sav     RFAAP  YGGAYKLG  VTVFGESAGAGSIASLLAMPSASGLF  FARTG  RLVQVLD
Sco     RFAPP  HGGGFVAG  VTLFGQSAGATVVGGVLATPEADGLF  FARTG  RATMRID
Bsu     RFKAP  HGGAFYLG  VTVFGESAGGMSIAALLAMPAAKGLF  FAKTG  RETVILD
Ccr     RFMPP  HGGAYNGG  VMLFGQSGGGAKIATLMAMPSAKGLF  FARTG  RQTMVFD
        ** .*  :**.:  *  * :**:*.*.  :  ::* * * ***  ** .*  * .  :*

C           I              II               III             IV         V
LR10    DGDFIDVDC  PILVLFHGLEGNAQSHY  GVSLGGNALMKWLGE  PTLLLMPLNDPF  GGHVGF
Rso     DGDFIDLDW  PLVVLFHGLEGDSRSHY  GISLGGNMLLRHLGE  PVLVLNARNDPF  GGHVGF
Neu     DGDFIDIDW  PLVIMLHGLEGSSQSHY  GVSLGGNALLKWLGE  PTLLINARNDPF  GGHVGF
Ath     DNGSVALDW  PILILLPGLTGGSQDSY  GWSLGGNILVNYLGQ  PLLCIQAANDPI  GGHLGW
Hsa     DGGQLLLDW  PIVLLLPGITGSSQETY  GISFGGILVLNHLAQ  PVLYLSAADDPF  GGHIGF
        *.. : :*   *::::: *: *.::. *  * *:**  ::. *.:  * * : . :**:  ***:*:

D                I               II         III           IV
LR12B   ALGGGAAKGFAHIGVIKVL  GTSAGSLVG  VRASSAVPGVFSP  VDGGLTHPVPA
Bma     ALGGGAARGFAHVGVLKAL  GTSAGSVVG  VRASCSVPSVFEP  VDGGLVSPVPA
Dar     ALGGGAARGFAHIGVIKML  GTSAGAVVG  VRASASVPGVFQP  VDGGLTSPIPV
Sth     ALGGGGARGAAHVGVLRAL  GTSSGAMVG  LRATTAMPCIVRP  VDGGVVHKVPV
Eca     ALGSGAAKGWAHIGVLNAL  GCSIGALVG  MRASCSMPGLLSP  VDGAVVNPVPV
Bth     ALSGGFIKGFAHLGVMQAL  GVSAGALAG  VAASCCMPVMFAP  VDGGLMMNLPV
        **..*  :* **:**:. *  * * *::.*  : *: .:* :. *  ***.:   :*.
```

Fig. 4. Conserved sequence blocks from multiple sequence alignment of lipolytic proteins from pond water metagenomic library. (A) LR5 # DQ077742 with members of non-heme peroxidases lipolytic proteins: Atu, *Agrobacterium tumefaciens* str. C58 # AAK89901; Bpy, *Burkholderia pyrrocinia* # P25026; Rle, *Rhizobium leguminosarum* bv. *trifolii* # AAL17798; Ppu, *Pseudomonas putida* # BAA86922; Pae, *P. aeruginosa* PAO1 # G83304; Rer, *Rhodococcus erythropolis* # O05691. (B) LR8 # DQ077745 with members of family Bacillus esterases lipolytic proteins: Sav, *Streptomyces avermitilis* # BAB69209; Sco, *S. coelicolor* A3(2) # CAC37455; Bsu, *Bacillus subtilis* # P37967; Ccr, *Caulobacter crescentus* CB15 # AAK22784. (C) LR10 # DQ077747 with members of family abh_upf0017 (Esther database): Rso, *R. solanacearum* # AL646060; Neu, *Nitrosomonas europaea* # BX321860; Ath, *A. thaliana* # AL049862; Hsa, *Homo sapiens* # AY033290. (D) LR12B # DQ077748 with members of patatin-like phospholipase proteins: Bma, *Burkholderia mallei* ATCC 23344 # AAU48139; Dar, *D. aromatica* RCB # ZP_00152189; Sth, *Symbiobacterium thermophilum* IAM 14863 # BAD41183; Eca, *Erwinia carotovora* SCRI1043 # CAG75234; Bth, *Bacteroides thetaiotaomicron* VPI-5482 # NP_809687. All residues are identical (*), there are conserved substitutions (:) and there are semiconserved substitutions (.) in that column in the alignment.

failed to find similar domain in other close matches of this protein from the databases.

Lipolytic protein of 343 amino acids was identified from 1622 bp DNA sequence obtained for the pLR7. It had maximum homology with putative hydrolase from *S. coelicolor* with 42% amino acid identity. This protein contained an α/β hydrolase conserved domain and belonged to family Moraxella lipase 3 of non-heme peroxidase superfamily of lipolytic proteins (Fig. 3D).

Identified protein from pLR8 showed maximum homology with Esterase 54 of *B. niacini* with 38% amino acid identity. This encoded protein of 458 amino acids belonged to Bacillus esterase family of carboxylesterase superfamily of lipolytic enzymes (Fig. 4B). Analysis in prosite database revealed carboxylesterase type B serine active site signature1 {F-[GR]-G-X(4)-[LIVM]-X-[LIV]-X-G-X-S-[STAG]-G} in amino acid sequence of encoded protein from pLR8.

An ORF of 227 amino acids from pLR9 in the lipolytic gene region matched with biotin biosynthesis protein, BioH of *X. axonopodis* with 47% amino acid identity. Comparison of encoded protein with other BioH proteins revealed truncation at N-terminal region of the encoded protein. Search for alternate start sites revealed the presence of a TTG and a GTG initiation codon at 60 and 96 bp upstream of the identified ATG codon. A putative RBS was identified upstream of the GTG codon indicating that the GTG codon might be the actual initiation codon for the lipolytic protein encoded from pLR9. The *bioH* gene from *E. coli* has been shown to possess esterase activity after the structure analysis revealed the presence of catalytic triad and active site similar to lipolytic enzymes [36].

The DNA sequence from the lipolytic gene region of pLR10 encoded a protein of 334 amino acids with best match to a predicted hydrolase of *D. aromatica* with 55% amino acid identity. Conserved domain search at NCBI and InterProScan revealed the presence of predicted α/β hydrolase domain in the encoded protein. No significant match was found in lipase engineering database for this protein. Search in Esther database revealed maximum match of this protein with members of uncharacterized α/β hydrolase protein family abh_upf0017. Protein members from bacteria, yeast, plants, and animals represent this uncharacterized family. Prosite database contains a signature pattern {D-X(8)-[GH]-[LFY]-X(4)-[DET]-[LY]-Y-X(3)-[ST]-X(7)-[IV]-X(2)-[PS]-X-[LIVM]-X-[LIVM]-X(3)-[DN]-D} for this family. Lipolytic protein from pLR10 contained this conserved pattern with only one amino acid difference. These proteins along with encoded protein from pLR10 contain the characteristic catalytic triad active site serine motif GXSXG, conserved aspartic acid, and histidine residues of lipolytic enzymes (Fig. 4C). These results indicate that encoded protein from pLR10 is a new member of yet uncharacterized α/β hydrolase protein family and members of this family are expected to possess lipolytic activity.

DNA sequence obtained from pLR12 contained two tandem ORFs similar to lipolytic genes. We obtained transposon mutants with no lipolytic activity located in both the identified probable lipolytic genes in pLR12; hence, it is difficult to predict if the activity loss is because of disruption of the individual gene or because of the polarity effect. As both the genes show significant similarity with lipolytic proteins, we assume that both the identified genes may possess lipolytic activity. One of the ORF encoded for a 301 amino acid long protein, which showed maximum homology to predicted hydrolase or acyltransferase of *R. gelatinosus* with 50% amino acid identity. The encoded protein belonged to family Moraxella lipase 3 of non-heme peroxidase superfamily (Fig. 3D). The second ORF encoded a protein of 287 amino acids, which was most similar to predicted esterase of *D. aromatica* with 56% amino acid identity. Conserved domain search revealed the presence of complete patatin and esterase domains. The patatin domain containing bacterial lipolytic proteins form a newly described probable bacterial lipolytic protein family [8]. This encoded protein was predicted to have a signal peptide and a transmembrane region from 7th to 29th amino acid at amino terminal.

Discussion

Majority of bacteria in environment remain unculturable by the routine laboratory cultivation methods. Culture-independent microbial diversity analysis revealed the presence of unculturable bacteria in most environmental samples. Half of the known phyla in eubacteria are not represented by any cultured member [37], indicating the vast phylogenetic diversity of these ‘‘yet to be cultured’ bacteria. Metagenomic libraries have been exploited for identification of novel genes for bioactive compounds and biocatalysts [11,12,14,38] avoiding the culture bias. Fresh water samples have been shown to possess high microbial diversity including members of unculturable clades. We have constructed a metagenomic library of pond water microbial assemblage containing about 532 Mb of community DNA. A part of this library was screened for isolation of 11 unique lipolytic clones, which is significantly higher in comparison to the lipolytic clones isolated from soil libraries [16,17]. Isolation of low number of functional lipolytic clones can be attributed to probable non-recognition of regulatory elements from unculturable bacteria in *E. coli* and reported difficulties in expression of lipolytic genes in heterologous hosts because of requirement of folding or transport partners [2] or because of toxicity of expressed lipolytic protein [19]. Isolation of higher number of functional lipolytic clones from our sample may be

because of our careful and deliberate attempt to select clones even with very low lipolytic activity, reasoning that many of the novel lipolytic genes may not express well in *E. coli* because of the above-mentioned factors. The other reason can be the difference in the composition of the bacterial assemblage of pond water sample in comparison to soil samples used in the other studies.

Significant G + C composition difference varying from 56% to 75% was observed in the DNA sequence of lipolytic gene containing region of the clones. In addition, the encoded proteins matched with 25–70% amino acid identity with proteins from bacteria belonging to proteobacterial, actinobacterial, and firmicute groups. This indicated that the DNA belonging to various phylotypes representing diversity of pond water microbial assemblage was cloned in the metagenomic library. This is in good correlation with other studies demonstrating expression of genes from various metagenomic samples in *E. coli* [11,12,14,15], indicating suitability of *E. coli* as a host for metagenomic studies. However, the use of other hosts is expected to increase the discovery rate of the functional clones.

Lipolytic clones identified from the pond water metagenomic library differed significantly in their ability to hydrolyze tributyrin in agar plates. There can be many possible reasons for this variable activity of the clones; one could be the difference in expression level of the different cloned genes. The G + C contents of the sequenced clones vary significantly and in many cases, we failed to identify the putative ribosome-binding site near to putative start site, which reflect that yet uncharacterized bacteria in pond water microbial assemblage differ in codon usage and regulatory elements from *E. coli*. In one of the clones, pLR5, a TTG and in another, pLR9, GTG codons were identified as putative initiation site. Alternate initiation codons TTG and GTG are known to generally result in lower expression than ATG initiation codon in *E. coli* [39]. Other possible reason could be the intracellular or membrane bound (LR2 and LR12B) nature of the encoded lipolytic proteins, which will not allow enzyme to diffuse in the plate and would only allow formation of hydrolysis zone near to the colony. Another reason could be the difference in the substrate preference of the encoded enzymes as each of the identified protein differed significantly from each other in amino acid sequence. Ten of the identified proteins belonged to seven known lipolytic protein families, one of the proteins matched with BioH and other with yet uncharacterized α/β hydrolase family. Members of different families of lipolytic enzymes have been shown to vary in their substrate preference. The substrate used in the plate assay may not be the most preferred substrate for the lipolytic enzymes produced by the clones isolated from pond water metagenomic library.

In this study and in a previous functional metagenomic study [18] lipolytic proteins similar to BioH were identified. Recently function for this protein of biotin biosynthesis operon was identified by structural genomics approach as an esterase [36]. Structural analysis of BioH protein from *E. coli* revealed the presence of a catalytic triad similar to lipolytic enzymes, which prompted study on the catalytic activity of this protein. Encoded protein from pLR10 showed high homology to proteins belonging to family abh_upf0017 in Esther database. Our results indicate that yet uncharacterized abh_upf0017 family members are expected to possess lipolytic activity. Identification of BioH like proteins and a novel lipolytic protein with high similarity to members of uncharacterized abh_upf0017 protein family reflects the potential of functional metagenomic studies in identification of function for many hypothetical and unknown function proteins even from the genome sequences of well-characterized culturable bacteria.

In conclusion, we have prepared a metagenomic library from pond water microbial assemblage containing more than 532 Mb of community DNA. Twelve unique lipolytic genes were identified with low similarity to known lipolytic proteins. Ten of the identified lipolytic proteins represented seven known lipolytic enzyme families. An encoded protein was most similar to recently identified esterase BioH. One novel lipolytic protein with low homology to known lipolytic proteins, belonging to uncharacterized α/β hydrolase family, was identified. They are expected to have different substrate specificity, which will be useful in developing them as biocatalysts for synthetic organic chemistry. Overexpression and purification of the proteins identified in this study are in progress to further characterize their substrate preference and biotechnological potential.

Acknowledgments

We thank Prof. Samir K. Brahmachari, Director, I.G.I.B. for encouragement and support. R.S. thanks Prof. Rup Lal for allowing use of his laboratory facilities in initial phase of this work. R.R. and A.G. acknowledge CSIR for pre-doctoral research fellowship. This work was supported by Grant SR/FTP/LS-277/2000 from Department of Science and Technology.

References

[1] R. Verger, Interfacial activation of lipases: facts and artifacts, Trends Biotechnol. 15 (1997) 32–38.
[2] K.E. Jaeger, T. Eggert, Lipases for biotechnology, Curr. Opin. Biotechnol. 13 (2002) 390–397.
[3] J.C. Moore, F.H. Arnold, Directed evolution of a *para*-nitrobenzyl esterase for aqueous-organic solvents, Nat. Biotechnol. 14 (1996) 458–467.
[4] U.T. Bornscheuer, J. Altenbuchner, H.H. Meyer, Directed evolution of an esterase for the stereoselective resolution of a key

intermediate in the synthesis of epothilones, Biotechnol. Bioeng. 58 (1998) 554–559.
[5] M. Fischer, J. Pleiss, The lipase engineering database: a navigation and analysis tool for protein families, Nucleic Acids Res. 31 (2003) 319–321.
[6] J.L. Arpigny, K.E. Jaeger, Bacterial lipolytic enzymes: classification and properties, Biochem. J. 343 (1999) 177–183.
[7] H.K. Kim, Y.J. Jung, W.C. Choi, H.S. Ryu, T.K. Oh, J.K. Lee, Sequence-based approach to finding functional lipases from microbial genome databases, FEMS Microbiol. Lett. 235 (2004) 349–355.
[8] S. Banerji, A. Flieger, Patatin-like proteins: a new family of lipolytic enzymes present in bacteria? Microbiology 150 (1999) 522–525.
[9] C. Upton, J.T. Buckley, A new family of lipolytic enzymes? Trends Biochem. Sci. 20 (1995) 178–179.
[10] R.I. Amann, W. Ludwig, K.H. Schleifer, Phylogenetic identification and in situ detection of individual microbial cells without cultivation, Microbiol. Rev. 59 (1995) 143–169.
[11] M.R. Rondon, P.R. August, A.D. Bettermann, S.F. Brady, T.H. Grossman, M.R. Liles, K.A. Loiacono, B.A. Lynch, I.A. MacNeil, C. Minor, C.L. Tiong, M. Gilman, M.S. Osburne, J. Clardy, J. Handelsman, R.M. Goodman, Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms, Appl. Environ. Microbiol. 66 (2000) 2541–2547.
[12] M.T. Cottrell, D.N. Wood, L. Yu, D.L. Kirchman, Selected chitinase genes in cultured and uncultured marine bacteria in the α- and γ-subclasses of the proteobacteria, Appl. Environ. Microbiol. 66 (2000) 1195–1201.
[13] J.K. Rhee, D.G. Ahn, Y.G. Kim, J.W. Oh, New thermophilic and thermostable esterase with sequence similarity to the hormone-sensitive lipase family, cloned from a metagenomic library, Appl. Environ. Microbiol. 71 (2005) 817–825.
[14] E.M. Gabor, E.J. de Vries, D.B. Janssen, Construction, characterization, and use of small-insert gene banks of DNA isolated from soil and enrichment cultures for the recovery of novel amidases, Environ. Microbiol. 6 (2004) 948–958.
[15] D.E. Robertson, J.A. Chaplin, G. DeSantis, M. Podar, M. Madden, E. Chi, T. Richardson, A. Milan, M. Miller, D.P. Weiner, K. Wong, J. McQuaid, B. Farwell, L.A. Preston, X. Tan, M.A. Snead, M. Keller, E. Mathur, P.L. Kretz, M.J. Burk, J.M. Short, Exploring nitrilase sequence space for enantioselective catalysis, Appl. Environ. Microbiol. 70 (2004) 2429–2436.
[16] A. Henne, R.A. Schmitz, M. Bomeke, G. Gottschalk, R. Daniel, Screening of environmental DNA libraries for the presence of genes conferring lipolytic activity on *Escherichia coli*, Appl. Environ. Microbiol. 66 (2000) 3113–3116.
[17] S.W. Lee, K. Won, H.K. Lim, J.C. Kim, G.J. Choi, K.Y. Cho, Screening for novel lipolytic enzymes from uncultured soil microorganisms, Appl. Microbiol. Biotechnol. 65 (2004) 720–726.
[18] H.C. Rees, S. Grant, B. Jones, W.D. Grant, S. Heaphy, Detecting cellulase and esterase enzyme activities encoded by novel genes present in environmental DNA libraries, Extremophiles 7 (2003) 415–421.
[19] P.J.L. Bell, A. Sunna, M.D. Gibbs, N.C. Curach, H. Nevalainen, P.L. Bergquist, Prospecting for novel lipase genes using PCR, Microbiology 148 (2002) 2283–2291.
[20] G. Zwart, B.C. Crump, M.P. Kamst-van Agterveld, F. Hagen, S.K. Han, Typical freshwater bacteria: an analysis of available 16S rRNA gene sequences from plankton of lakes and rivers, Aquat. Microb. Ecol. 28 (2002) 141–155.
[21] F.O. Glockner, E. Zaichikov, N. Belkova, L. Denissova, T. Pernthaler, R. Amann, Comparative 16S rRNA analysis of lake bacterioplankton reveals globally distributed phylogenetic clusters including an abundant group of Actinobacteria, Appl. Environ. Microbiol. 66 (2000) 5053–5065.
[22] J. Zhou, M.A. Bruns, J.M. Tiedje, DNA recovery from soils of diverse composition, Appl. Environ. Microbiol. 62 (1996) 316–322.
[23] F.M. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith, K. Struhl (Eds.), Current Protocols in Molecular Biology, John Wiley & Sons, New York, 2001.
[24] S. Wilhelm, J. Tommassen, K.E. Jaeger, A novel lipolytic enzyme located in the outer membrane of *Pseudomonas aeruginosa*, J. Bacteriol. 181 (1999) 6977–6986.
[25] J. Sambrook, D.W. Russell, third ed.Molecular Cloning: A Laboratory Manual, vols. 1–3, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 2001.
[26] D.L. Wheeler, D.M. Church, S. Federhen, A.E. Lash, T.L. Madden, J.U. Pontius, G.D. Schuler, L.M. Schriml, E. Sequeira, T.A. Tatusova, L. Wagner, Database resources of the National Center for Biotechnology, Nucleic Acids Res. 31 (2003) 28–33.
[27] S.F. Altschul, T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, D.J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acids Res. 25 (1997) 3389–3402.
[28] T. Hotelier, L. Renault, X. Cousin, V. Negre, P. Marchot, A. Chatonnet, ESTHER, the database of the a/b-hydrolase fold superfamily of proteins, Nucleic Acids Res. 32 (2004) D145–D147.
[29] A. Marchler-Bauer, S.H. Bryant, CD-Search: protein domain annotations on the fly, Nucleic Acids Res. 32 (2004) W327–W331.
[30] E.M. Zdobnov, R. Apweiler, InterProScan—an integration platform for the signature-recognition methods in InterPro, Bioinformatics 17 (2001) 847–848.
[31] N. Hulo, C.J.A. Sigrist, V. Le Saux, P.S. Langendijk-Genevaux, L. Bordoli, A. Gattiker, E. De Castro, P. Bucher, A. Bairoch, Recent improvements to the PROSITE database, Nucleic Acids Res. 32 (2004) D134–D137.
[32] J.D. Thompson, D.G. Higgins, T.J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, Nucleic Acids Res. 22 (1994) 4673–4680.
[33] J.D. Bendtsen, H. Nielsen, G.V. Heijne, S. Brunak, Improved prediction of signal peptides: SignalP 3.0, J. Mol. Biol. 340 (2004) 783–795.
[34] G.E. Tusnady, I. Simon, The HMMTOP transmembrane topology prediction server, Bioinformatics 9 (2001) 849–850.
[35] G. Kouker, K.E. Jaeger, Specific and sensitive plate assay for bacterial lipases, Appl. Environ. Microbiol. 53 (1987) 211–213.
[36] R. Sanishvili, A.F. Yakunin, R.A. Laskowski, T. Skarina, E. Evdokimova, A.D. Kirby, G.A. Lajoie, J.M. Thornton, C.H. Arrowsmith, A. Savchenko, A. Joachimiak, A.M. Edwards, Integrating structure, bioinformatics, and enzymology to discover function, J. Biol. Chem. 278 (2003) 26039–26045.
[37] P.D. Schloss, J. Handelsman, Status of the microbial census, Microbiol. Mol. Biol. Rev. 68 (2004) 686–691.
[38] S. Courtois, C.M. Cappellano, M. Ball, F.X. Francou, P. Normand, G. Helynck, A. Martinez, S.J. Kolvek, J. Hopke, M.S. Osburne, P.R. August, R. Nalin, M. Guerineau, P. Jeannin, P. Simonet, J.L. Pernodet, Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products, Appl. Environ. Microbiol. 69 (2003) 49–55.
[39] C.M. Stenstrom, E. Holmgren, L.A. Isaksson, Cooperative effects by the initiation codon and its flanking regions on translation initiation, Gene 273 (2001) 259–265.