Protein Engineering vol. 16 no. 5 pp. 357–364, 2003
DOI: 10.1093/protein/gzg043

A systematic approach for yielding a potential pool of enzymes: practical case for chiral resolution of (*R*,*S*)-ketoprofen ethyl ester

Ji-Youn Kim[1], Gi-Sub Choi[1], In-Su Jung[2], Yeon-Woo Ryu[1] and Geun-Joong Kim[2,3]

[1]Department of Molecular Science and Technology, College of Engineering, Ajou University, San 5, Woncheon-dong, Paldal-gu, Suwon 442-749 and [2]Institute of Biotechnological Industry, College of Engineering, Inha University, 253 Yonghyun-dong, Nam-gu, Incheon 402-751, Korea

[3]To whom correspondence should be addressed.
E-mail: geunkim@inha.ac.kr

A systematic approach for the selection of potential biocatalysts from a natural source was developed and then a practical application was addressed. The approach that involves systematically combined conventional screening methods and current tools comprises the following consecutive steps: strain enrichment for activity screening, identification of positive strains, choosing whole genome-sequenced strains as candidates, gathering information about responsible enzymes, bioinformatic analyses and gene mining, probing genetic molecules and then functional expression. The target compound (*R*,*S*)-ketoprofen ethyl ester was to be resolved into an enantiomer, and a potential esterase from *Pseudomonas fluorescens* KCTC 1767 was prepared by the proposed procedure. The enzyme had a high activity and also strict selectivity for the enantiomer (*S*)-ketoprofen and was suitable therefore as a biocatalyst for practical use. The result achieved by the combined approach could not easily be obtained using other approaches with typical procedures. Hence the approach proposed here should be of considerable use for the screening of potential enzymes, particularly for enzymes with desired activity to unnatural substrates, from conditionally expressed and/or repressed proteins that are distributed widely in natural pools under normal conditions.

Keywords: bioinformatics/chiral resolution/esterase/gene mining/(*S*)-ketoprofen

Introduction

Biological catalysts, mainly enzymes, in general, are attractive for industrial purposes because they are efficient and selective in the mediated reactions in which they are utilized (Bornscheuer and Kazlauskas, 1999; Jaeger *et al*., 1999). Enzymes are functionally precise and have a good performance for most reactions found naturally or exploited industrially. Moreover, ecological niches have been providing diverse microorganisms that adapt well to their functions through the evolutionary process and thus abundant enzyme pools exist securely in nature and are available for the screening of enzymes with desired properties, whether or not the activity has been screened to date (Dalboge and Lange, 1998). Therefore, there are continuing and also increasing demands for the screening of new enzymes with high activity and selectivity for specific chemical compounds (Ogawa and Shimizu, 1999).

Non-steroidal anti-inflammatory drugs (NSAIDs) have been widely used for alleviating pain and inflammation associated with tissue injury (Hayball, 1996). Among them, ketoprofen [(*R*,*S*)-2-(3-benzoylphenyl)propionic acid], as an *in vitro* inhibitor of prostaglandin synthesis, is one of the most prevalent anti-inflammatory agents belonging to the class of 2-arylpropionic acids (Mauleon *et al*., 1996; Patal, 2000). Since ketoprofen has been produced by chemical synthesis and thus sold as a mixture of stereoisomers, much effort is currently being devoted to enzymatic synthesis of optically pure ketoprofen for pharmacological use, because the activity is mainly exerted by the *S*-enantiomer (Caldwell *et al*., 1988). In this context, with their abundance and great versatility in mediated reactions, esterases have recently been considered as a possible candidate for the chiral resolution of ketoprofen (Shen *et al*., 2001; Bornscheuer, 2002; Kim *et al*., 2002).

We have recently screened various ecological niches and found a strain that expresses a novel esterase with relatively high activity and enantioselectivity (Kim *et al*., 2002). Although the results had partly seemed promising for practical applications, further attempts to obtain more potential enzymes failed, because the known information and also screening procedures were restricted to the whole cell level, mainly owing to the absence of established genetic data and sequence information. Therefore, we have designed a protocol to extend the scope for the screening of potential enzymes from natural sources.

In this paper, we present a systematic approach to a screening procedure for the isolation of potential enzymes from the natural pool of strains by combining pre-existing tools. The principle and practical application of this systematic approach are demonstrated by employing (*R*,*S*)-ketoprofen ethyl ester as a target compound for esterase-mediated chiral resolution.

Materials and methods

Chemicals and enzymes

Fast Blue RR, (*R*,*S*)-ketoprofen, α-naphthyl acetate, tributyrin, Triton X-100 and ethoxyethanol were purchased from Sigma, acrylamide stock (30%) and protein assay solutions from Bio-Rad and agarose from Promega. Restriction enzymes, *Taq* DNA polymerase and T4 DNA ligase were obtained from KOSCO (Korea) and Vent DNA polymerase from New England Biolabs. (*R*,*S*)-Ketoprofen ethyl ester was prepared using a general procedure for esterification (Kim *et al*., 2000).

Strain enrichment, selection and identification

Hundreds of samples were collected from various ecological niches, such as compost, forest, contaminated soils and sewage sludge, and suspended in a phosphate buffer (0.85% saline solution) or a previously formulated TS medium without

tributyrin (Kim *et al*., 2002). Enrichment for the activity screening was also performed by using TS medium, as described previously (Kim *et al*., 2002). The screened colonies were then inoculated into 50 ml of TS medium with or without 10 mM ketoprofen ethyl ester. After cultivation at 30°C for 24 h, the cells were harvested and then confirmed for activity to *rac*-ketoprofen ethyl ester. The strains showing a high ester-hydrolyzing activity were finally screened and systematically identified by general procedures from Bergey's Manual of Systematic Bacteriology and 16S rRNA sequencing, and further classified by morphological and genetic traits.

Cell growth and fractionation of crude extracts

After sorting the positive strains, the strain pool of *Pseudomonas* was cultivated at 30°C in 500 ml baffled flasks containing 100 ml of culture medium supplemented or not with ester derivatives as an inducer. Inoculating cells were prepared by pre-culture of a well-isolated colony in TS medium (10 ml) for 24 h. During the cultivation, an aliquot of culture broth was sampled and analyzed for cell growth and enzyme activity.

To obtain information at the protein level, crude extracts were fractionated to prepare the fraction with an activity to *rac*-ketoprofen ethyl ester. *Pseudomonas* cells from 0.5 l of culture broth were harvested and resuspended in a buffer A (20 mM Tris–HCl, pH 8.0) containing protease inhibitor cocktail solution (Sigma) and disrupted using a sonicator. Cell debris was removed by centrifugation at 14 000 *g* for 30 min and the supernatant was treated with protamine sulfate (2%). After centrifugation, the supernatant was desalted with a prepacked column (Sephadex G-25) and then loaded on to an ion-exchange column (Resource Q, 6 ml) using an FPLC system (Pharmacia). The column was washed with 10 volumes of buffer A and eluted with a linear gradient of 0–0.5 M NaCl. The active fractions were pooled and then stored for further characterization.

Molecular mass, specific activity and enantioselectivity

The molecular masses and oligomeric structures of probable esterases in active fractions were determined on an FPLC system with a gel filtration column (Superdex-75). The flow rate of the mobile phase, containing 20 mM Tris–HCl and 150 mM NaCl, was 0.5 ml/min. A molecular mass calibration curve was established by plotting the elution volumes of protein markers (Sigma) versus the known molecular masses on semi-logarithmic paper. Aliquots of eluted fractions were analyzed for the activity to *rac*-ketoprofen ethyl ester.

To determine the specific activity and enantioselectivity, the activity on the ketoprofen ethyl ester was determined at 30°C for 30–60 min with the fractionated enzyme (30–50 mg) in 3 ml of reaction mixture containing 50 mM Tris–HCl (pH 8.5), 0.3% Triton X-100 and 5 mM substrate. The reaction was stopped by the addition of four volumes of ethanol (>99%) and then analyzed for the conversion yield and chiral selectivity.

Native gel electrophoresis and activity staining

For activity staining on native PAGE, protein samples were mixed with 0.2 vol. of a native sample buffer and resolved on a 12% gel under non-reducing conditions (Kim *et al*., 2002). After gel electrophoresis, the separating gel was washed twice with 20 ml of 50 mM Tris–HCl buffer and then soaked in the same buffer (100 ml) containing 4 mg of α-naphthyl acetate dissolved in 0.5 ml of ethoxyethanol. The band corresponding to the active enzyme was visualized by the addition of Fast Blue RR (2 mg/ml) and then sliced. The activity was further confirmed with the sliced gel by using *rac*-ketoprofen ethyl ester as a substrate.

Protein database search and sequence alignment

With the information obtained from the fractionated crude extracts, enzymes and their sequences were searched for in databanks, including the complete genome sequence of *Pseudomonas aeruginosa* PA01 (Stover *et al*., 2000) and then analyzed further using the BLAST network service at the National Center for Biotechnological Information (NCBI). To access closely to the best candidate genes or proteins, valuable data, such as N-terminal amino acid sequence, ORF size, quaternary structure, apparent PI and substrate spectrum, were fully considered for gene mining. Primarily selected probable sequences were compared manually or automatically on annotated sequences of an esterase pool from various genomes and the protein data bank. The resulting sequences were aligned by hierarchical clustering of the individual sequences based on the pairwise similarity scores. The conserved patterns of amino acid sequences in related enzymes were analyzed with the Clustal W program (Thompson *et al*., 1994) and then confirmed by visual inspection. The distribution of the secondary structural elements in these sequences was obtained by using a previous algorithm (Cuff and Barton, 2000).

Probing the related genes by low-stringency PCR

The amplification of probable genes was performed by PCR with two wobble primers, N-terminal ESTF, 5′-A(G)TGCAG(C)ATTCA(G)G(A)GGT(C)CAT(C)TAC(T)-GAA(G,C)-3′, and C-terminal primer ESTR, 5′-TTACAG-ACAA(C)C(G)C(T)G(C)G(C)CCA(G)A(G)TAT(C)(T,A)TC-C(G)-3′, which were designed based on terminal sequences of systematically mined esterase genes. Chromosomal DNA was isolated from each strain of *Pseudomonas* by using a kit (Promega). The partially digested or intact genomic DNA (2.5–20 ng) was used as the template for PCR (1 cycle, 94°C, 7 min; 40 cycles, 94°C, 1 min, 37.5 ± 0.4°C/cycle, 1 min, 72°C, 90 s; 1 cycle, 72°C, 7 min). To expand the amplifying sequence, the stringency at the annealing step was further modulated (Kim *et al*., 2001).

Cloning and selection of a potential esterase

The amplified DNA fragments under low-stringency conditions were reamplified by PCR with two primers, ESTF1 and ESTR1, which introduced the *Bam*HI and *Pst*I sites into the N- and C-terminal primer, respectively. For rapid purification and high-level expression, the reamplified genes were subcloned into the *Bam*HI and *Pst*I sites of a series vector pQE 30, 31 and 32 and thus expressed as poly-His-tag fusion proteins at their N-termini using a different reading frame, according to the general procedure of the manufacturer.

The positive clones were screened from transformants by activity staining using an overlaid soft agar (0.6%) containing Fast Blue RR (15 mg/ml) and α-naphthyl acetate (45 mg/ml). α-Naphthyl acetate was first dissolved in ethoxyethanol and then added to soft agar solution. The positive clones rapidly (<5 min) developed a deep brown color around the colony, as reported previously (Kim *et al*., 2002).

Enzyme purification and characterization

Escherichia coli cells expressing the fusion protein were grown in 200 ml of LB medium at 30°C to an $OD_{600\ nm}$ of 0.4–0.5 and then induced with 0.2 mM IPTG for 2 h. After centrifugation at 10 000 *g* for 10 min, the cells were resuspended in 10 ml of phosphate buffer (50 mM, pH 8.0) containing 10 mM

Table I. Ketoprofen ethyl ester hydrolyzing activities of *Pseudomonas* strains

Strain	TS			LB		
	Conversion (%)	ee_p[a] (%)	E[b]	Conversion (%)	ee_p (%)	E
P.aeruginosa						
KCTC1636	<1	34	2	2	82	10.3
KCTC2450	15	97	77	34	95	63.4
P.fluorescens						
KCTC1767	2	–9	0.8	4	57	3.7
KCTC2344	9	3	1.1	11	34	2.1
P.putida						
KCTC1642	4	–100	0	5	52	3.3
Pseudomonas sp.						
S34	35	97	110	42	95	80.5
HS12	21	93	35	N.D.[c]	–	–

[a]Enantiomeric excess, ee_p, was calculated based on the following equation: ee_p = [(*S*)-ketoprofen] – [(*R*)-ketoprofen]/[(*S*)-ketoprofen] + [(*R*)-ketoprofen] × 100.
[b]Enantiomeric ratio, $E = \ln[1 - c(1 + ee)]/\ln[1 - c(1 - ee)]$.
[c]Not determined.

imidazole. The suspended cells were sonicated and then centrifuged at 18 000 *g* for 30 min. The resulting supernatant was loaded on to Ni-NTA resin (Qiagen) and the bound protein was eluted with a buffer containing 250 mM imidazole.

The enzyme properties were analyzed in terms of molecular mass, specific activity, substrate spectrum, enantioselectivity and conversion yield, by using the purified enzyme, according to the same procedures as above and from elsewhere (Fernandez *et al*., 2000; Kim *et al*., 2002). All experiments were conducted in duplicate and mean values were calculated.

Analyses

The concentrations of (*R*,*S*)-ketoprofen ethyl ester and (*R*)- and (*S*)-ketoprofen were determined by HPLC (Waters). The column and mobile phase used were Chirex Phase 3005 (Phenomenex) and methanol containing 30 mM ammonium acetate, respectively. At a constant flow rate (0.8 ml/min), the eluate was monitored at 254 nm. One unit of esterase activity was defined as the amount of enzyme producing 1 mmol of ketoprofen from the corresponding ethyl ester per minute under the specified conditions. Protein concentration was measured by using a protein assay solution (Bio-Rad).

Results and discussion

A systematic approach for the selection of a potential enzyme pool

Step 1. Strain enrichment and activity screening. By considering the previous results that established a screening procedure for the strains expressing an ester-hydrolyzing enzyme, the activity screening was repeated to broaden the strain pools (Kim *et al*., 2002). From various environmental samples, thousands of strains were first enriched and then analyzed for activity by using a TS agar plate or a selective plate overlaid with soft agar (0.6%) containing Fast Blue RR and α-naphthyl acetate as an activity indicator. Based on the developed clear zone around colonies, ~150 strains were readily detected and further confirmed the activity to (*R*,*S*)-ketoprofen ethyl ester under favorable conditions for cell growth and enzyme assay. The strains showing a high growth rate and distinct clear zone were compared again for their potential for activity and enantioselectivity in a solid or solution culture and then finally the strains that had a strict selectivity to an *S*- or *R*-enantiomer form werescreened. As a result, 24 strains were chosen as possible candidates for further analyses.

Step 2. Strain identification and sorting. The next step was a further comparison with respect to the phenotypic and genotypic characters for strain classification. Among the screened isolates, the strain pool (nine out of 24 strains) was motile, rod-shaped, Gram-negative, no endospore-forming, pyruvate-utilizing and aerobic microorganisms. All strains showed both catalase- and oxidase-positive reactions and either fluorescence or not. They also utilized glucose, fructose, galactose and citrate as carbon sources. The optimal growth temperature was 25–30°C and they could not grow at 40°C. The G + C content of the genomic DNA was in the range 52–66%. The 16S rRNA sequence analyses revealed a fairly high homology (>96%) to the typical species of *Pseudomonas*, including *P.fluorescens*, *putida* and *aeruginosa*. The level of 16S rRNA identity, therefore, together with physiological and taxonomic properties, strongly suggested that the nine strains might be related to one of the typical strains of *Pseudomonas*. No distinct differences between the analyzed and documented characters of the genus were found as an ambiguous trait. Therefore, these strains were taxonomically identified as *Pseudomonas* strains according to Bergey's Manual of Systematic Bacteriology. The results were also confirmed by systematic approaches in the Korea Collection for Type Culture (KCTC). The remaining strains (15 out of 24) that had a relatively low activity were taxonomically identified as various strains including *Bacillus*, *Mycobacterium*, *Streptococcus* and *Rhodococcus* (data not shown).

Step 3. Data gathering from the strain pool of identified genus. Although the activity from *Pseudomonas* on the *rac*-ketoprofen ethyl ester was not fully established to date, the screening results provided the possibility that most *Pseudomonas* strains could hydrolyze the ethyl ester in a stereospecific manner, which was also supported by our previous work (Kim *et al*., 2002). A further attempt to confirm the hydrolyzing activity on the *rac*-ketoprofen ethyl ester was carried out by employing various species of *Pseudomonas*. For this purpose, two *fluorescens* (KCTC1767 and 2344), two *aeruginosa* (KCTC1636 and 2450) and a *putida* (KCTC1642) were randomly chosen from the KCTC as representative strains. The two isolates screened, *Pseudomonas* sp. S34 (Kim *et al*., 2002) and HS12 (Park and Kim, 2000), were also employed as possible sources for the activity. Among these strains, only one case (S34) has been reported to possess an activity to *rac*-ketoprofen ethyl ester.

The seven strains were first analyzed as possible sources for potential esterases by using a selective TS plate. All strains tested were well grown and developed a clear zone distinctly, although the colony size and clear zone were somewhat different. Further experimental results that exhibited the *rac*-ketoprofen ethyl ester hydrolyzing activity to the *R*- or *S*-enantiomer or both, strongly suggested the presence of an esterase in these strains. As shown in Table I, when the cells were grown in TS and LB media, different enantioselectivity and conversion yield were observed, with a similar trend in the relative activity. A relatively high activity and enantioselectivity were found in the three strains S34, KCTC2450 and HS12. In contrast, two strains, KCTC1636 and 1767, revealed a minor activity to ketoprofen ethyl ester. Two strains,

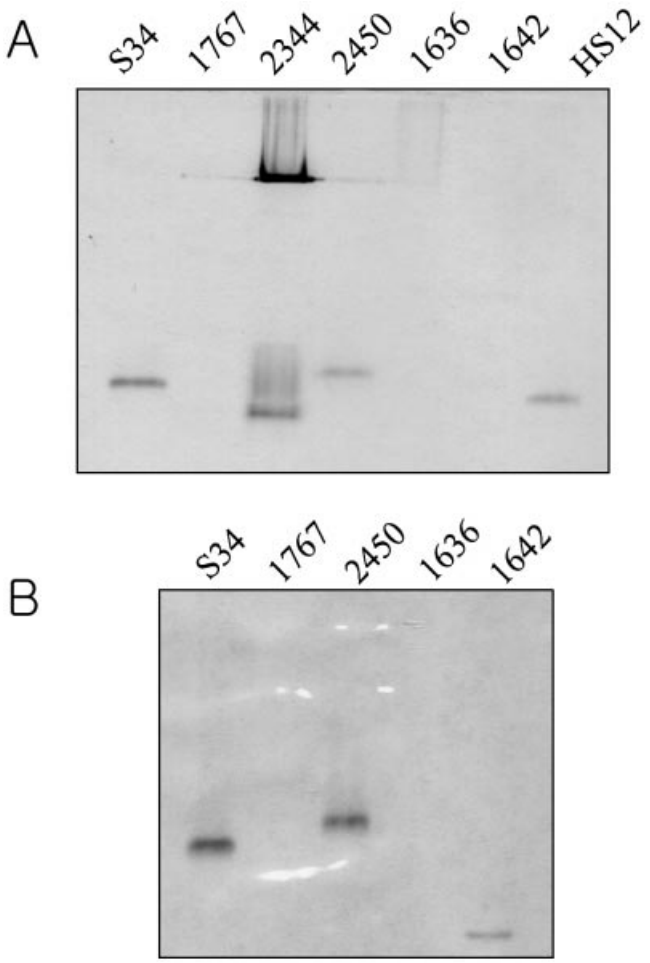

Fig. 1. Detection of ester-hydrolyzing activity from a strain pool of *Pseudomonas* by activity staining. Crude enzyme solutions (35–50 μg) were prepared from grown cells at 30°C in (**A**) TS and (**B**) LB media and loaded on to a lane of native PAGE. The activity staining was performed as described in Materials and methods.

KCTC2450 and S34, fully retained the enantioselectivity to the *S*-enantiomer, whereas the enantioselectivity in other strains was partly affected or reversed by the culture conditions. Other factors, such as temperature, inducers, carbon and nitrogen sources, that probably affected the cell growth and enzyme production, did not improve the results significantly.

For further analyses, crude extracts of the seven strains were prepared and then analyzed for activity staining on native PAGE. As expected, there was a strong correlation between the activity using the whole cell and crude enzymes (Figure 1). Cell extracts of the strains KCTC1636, 1642 and 1767 showed a negligible or no activity band under our conditions. Although a high concentration of protein was loaded (>200 mg), the activity was as low as detected in whole cell assays. These results provided the possibility that the enzymes might share a similar substrate spectrum, because the activity to ketoprofen ethyl ester and α-naphthyl acetate was well correlated, as shown in Table I and Figure 1. When the general criteria that distinguished esterases from lipases were considered (Bornscheuer, 2002), most of the enzymes involved here were considered to be esterases, based on primary screening on a tributyrin-supplemented selective plate, high activities on triglycerides composed of short-chain fatty acids (<C_6) and favorable reactions at alkaline pH, 8.0–9.5 (data not shown).

More information about the quaternary structures and molecular masses was partly obtained from activity staining, as shown in Figure 1. The developing rates and patterns on the native gel supported the enzymes as being a protein smaller than 50 kDa. These possibilities were also supported by the calculated molecular masses of 40–45 kDa, from the eluted fractions of gel filtration column chromatography. Therefore, there were at least two possible structures of enzymes, a monomer of 40–50 kDa or a dimer of two monomers (~20 kDa). No active fractions corresponding to a higher molecular mass or oligomeric structure appeared under our analytical conditions. The active fractions of each crude extract showed a similar level of activity as detected in the whole cell enzymes.

Step 4. Bioinformatic analyses and gene(s) mining. Gene mining using only preliminary biochemical and taxonomic data is partly troubleshooting. Fortunately, however, these data are valuable for a typical sequence that is predicted as an open reading frame from whole genome sequencing. In this context, data obtained from probable esterases could be sufficient to mine a sequence, at least to deduce a representative from the complete genome sequence of *P.aeruginosa* PA01 (Stover *et al*., 2000). In order to find candidate genes encoding an esterase, the whole genome sequence of PA01 was analyzed and then compared with data obtained above, yielding a sequence (NP_249738) as a possible one for the ketoprofen ethyl ester-hydrolyzing enzyme. In addition, the search results in protein data banks revealed a subfamily comprising a related set of esterases (AF228666, JC2091, A44832 and 2006221A), all of which also originated from the identical genus *Pseudomonas*. The first one is still uncharacterized to date and the last four enzymes have already been identified as a (carboxyl)esterase (McKay *et al*., 1992; Kim *et al*., 1994). However, not all cases have been reported on the activity to ketoprofen ethyl ester.

In accordance with the properties of crude enzymes, all mined enzymes were composed of a similar number of amino acid residues (377–392) and oligomeric structure (monomer). From the reported cases (McKay *et al*., 1992; Kim *et al*., 1994), the analyzed properties for optimal pH and substrate spectrum were also found to be similar to those of the crude extracts. The superimposable features in the nucleotide and amino acid sequences further supported the assumption. Consistent with these views, the N-terminal amino acid sequence (VQIQGHY–) of the esterase from S34 also supported the contention that the mined enzymes might be able to hydrolyze *rac*-ketoprofen ethyl ester, because the sequence was also found in the mined enzymes.

Step 5. PCR amplification for functional expression. From a further analysis of the mined sequences at the molecular level, the nucleotide and amino acid sequence identities at their N- and C-termini were also found to be rigidly conserved and thus sufficient to amplify the whole ORF from related chromosomes by using the degenerated primers (Figure 2). To verify and access the related genes, PCR amplifications were carried out using the chromosomal DNA of seven strains and a set of degenerated primers. Because the reaction yielded a smeared or no amplified DNA under normal conditions, low-stringent PCRs were selected to enhance the amplification of related genes (Kim *et al*., 2001; Yuen *et al*., 2001). Under these relaxed conditions, the expected size of DNA fragments

N-Terminal regions

```
JC2091           1 MQIQGHYELQFEAVREAFA 19
2006221A         1 MQIQGHYELQFEAVREAFA 19
A44832           9 VQIQGHYELQFEAIREAFA 27
AF228666         1 MQIQGHYELQFEAVREAFA 19
NP_249738       12 MTPQGHCDSRFAPLAEAFA 30
```

C-Terminal regions

```
JC2091         363 PYVLMDPRAQKLVGILAGCL 382
2006221A       363 PYVLMDPRAQKLVGILAGCL 382
A44832         370 PYVLMDPRAQKLVGILAGCL 389
AF228666       358 PYVLMDPRAQKLVGILAGCL 377
NP_249738      373 PYVLMDPRAQQLARVAGECL 392
```

Fig. 2. The conserved amino acid sequences in the N- and C-terminal regions of the mined esterase family enzymes. All of the enzymes originated from a strain of *Pseudomonas*, such as *P.fluorescens* (JC2091, 2006221A, AF228666), *Pseudomonas* sp. (A44832) and *P.aeruginosa* (NP_249738).

(~1.1–1.3 kb) were amplified by PCR from the chromosomal DNA of KCTC1642, 1767, 2344, 2450 and S34 (data not shown). When using two chromosomes (HS12 and KCTC1636) as templates, the reactions were inefficient and thus resulted in a minor or no band corresponding to the expected size.

To identify the amplified genes, the resulting fragments were reamplified with two primers, ESTF1 and ESTR1, and then subcloned into the expression vector pQE30. The transformed *E.coli* cells with these constructs were primarily screened for their activity on the TS plate and then further compared using whole cell enzyme. Unexpectedly, the clone with the amplified gene from KCTC1767 exhibited the highest activity, even under non-induced conditions. When induced at 30°C for 2 h, the specific activity of this clone was much higher (>11-fold) than the clones with genes from any other strains. It was noteworthy that the wild-type KCTC1767 revealed a minor and thus negligible activity in both whole cell and crude enzymes when compared with other strains (S34, HS12 and KCTC2450). To avoid the risk of causing different reading frames by a low-stringent PCR with degenerated primers, the amplified genes were also cloned and expressed in other vectors, pQE31 and 32, that utilized a different reading frame, but the results did not result in protein expression.

Characterization of a potential esterase

The crude extracts of *E.coli* cells harboring the gene from KCTC1767 showed a distinct protein band (>7% of total cell protein), corresponding to the expected size when induced with 0.25 mM IPTG for 2 h (Figure 3A). The enzyme was mainly expressed in the soluble fraction, while a minor portion was detected in the insoluble fraction (<5%). With the crude extracts of induced cells, the enzyme was easily purified to apparent homogeneity by using an affinity resin, Ni-NTA, in a single step (Figure 3A).

The enzyme was found to be a monomer with an apparent molecular mass of 41–43 kDa, calculated both from the elution volume of gel filtration chromatography and from the mobility on SDS–PAGE. Concerning the substrate specificity, the purified enzyme showed a relatively high activity to *p*-nitrophenyl acetate (5040 units/mg protein) and butyrate

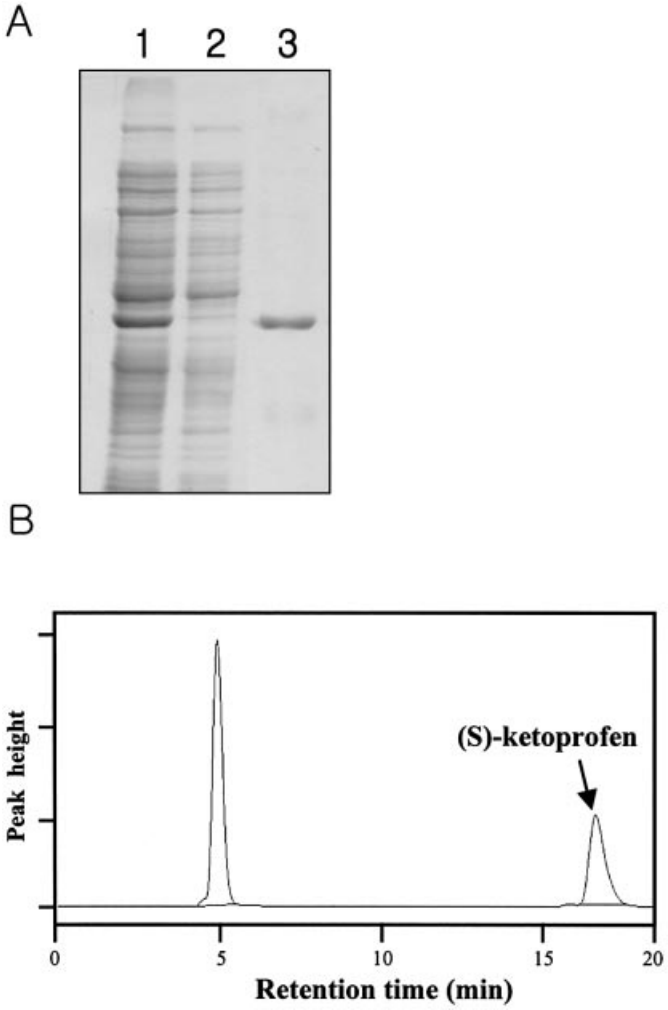

Fig. 3. Purification and enantioselectivity of an esterase from KCTC1767. (**A**) SDS–PAGE analysis of the affinity-purified enzyme. Protein samples were loaded on to a 12% polyacrylamide gel under denaturing conditions. Lane 1, soluble fraction of crude extract; lane 2, insoluble fraction of crude extract; lane 3, purified enzyme. (**B**) HPLC chromatogram of enzymatically resolved (*S*)-ketoprofen. The reaction product of the purified enzyme was analyzed on a chiral column and the retention time was compared with those of authentic samples: *rac*-ketoprofen ethyl ester, 4.8 min; (*R*)-ketoprofen, 14.3 min; (*S*)-ketoprofen, 15.7 min.

(2600 units), while the activities to *p*-nitrophenyl palmitate and caprylate were determined as about 7 and 2% of that to *p*-nitrophenyl acetate, respectively, partly supporting the enzyme being an esterase. There was negligible activity on *p*-nitrophenyl decanoate. Under standard assay conditions (pH 8.5 and 30°C), the esterase showed the highest activity towards ketoprofen ethyl ester (27 000 units), with a strict selectivity to the *S*-enantiomer (Figure 3B).

To compare the primary structure with other mined sequences, the nucleotide sequence of the PCR-amplified gene from KCTC1767 was completely determined on both strands. Analysis of the DNA sequence showed a complete ORF (1146 bp) of 381 amino acid residues with a calculated molecular mass of 41 kDa (Figure 4). As expected, its amino acid sequence revealed a close identity to the mined sequences used here for primer design. As reported previously in the related genes, the ORF did not utilize ATG but GTG as the start codon (Gold *et al*., 1981; Kim *et al*., 1994) and ended with

```
   1 GTG CAG ATT CAG GGT CAT TAC GAA CTT CAG TTC GAA GCG GTA CGT GAA GCG TTC GCC GCC    60
   1 V   Q   I   Q   G   H   Y   E   L   Q   F   E   A   V   R   E   A   F   A   A    20

  61 TTG TTC GAT GAC CCG CAG GAG CGT GGT GCC GGG CTT TGC ATC CAG ATC GGC GGG GAA ACG   120
  21 L   F   D   D   P   Q   E   R   G   A   G   L   C   I   Q   I   G   G   E   T    40

 121 GTA GTC GAC CTG TGG GCC GGT ACC GCC GAC AAG GAT GGT GCC GAG GCC TGG CAC AGC GAC   180
  41 V   V   D   L   W   A   G   T   A   D   K   D   G   A   E   A   W   H   S   D    60

 181 ACC CTC GTC AAC CTG TTC TCG TGC ACC AAG ACG TTT ACT GCG GTC ACG GCC CTG CAA TTG   240
  61 T   L   V   N   L   F   S   C   T   K   T   F   T   A   V   T   A   L   Q   L    80

 241 GTC GCC GAA GGC AAG CTG AAG CTG GAT GCG CCA GTG GCC GAT TAT TGG CCG GCG TTT GCG   300
  81 V   A   E   G   K   L   K   L   D   A   P   V   A   D   Y   W   P   A   F   A   100

 301 GCG GCG GGG AAG GAA ACC ATC ACC CTG CGC CAG TTG CTC TGC CAC CAG GCC GGA TTG CTG   360
 101 A   A   G   K   E   T   I   T   L   R   Q   L   L   C   H   Q   A   G   L   L   120

 361 GCG ATC CGC GAA ATG CTG CCC GCC GAA GCG CTA TAC GAC TGG CAA CGC ATG GTC GAC ACC   420
 121 A   I   R   E   M   L   P   A   E   A   L   Y   D   W   Q   R   M   V   D   T   140

 421 CTG GCG GCC GAA GCG CCG TGG TGG ACA CCG GGC CAA GGC CAT GGC TAC GAG GCA ATT ACC   480
 141 L   A   A   E   A   P   W   W   T   P   G   Q   G   H   G   Y   E   A   I   T   160

 481 TAT GGT TGG CTG GTC GGC GAA TTG CTG CGC CGT GCC GAC GGG CGC GGC CCG GGT GAA TCC   540
 161 Y   G   W   L   V   G   E   L   L   R   R   A   D   G   R   G   P   G   E   S   180

 541 ATC GTG GCG CGG GTT GCG CGG CCG TTG GGA CTG GAC TTC CAT GTA GGG CTG GCG GAT GAA   600
 181 I   V   A   R   V   A   R   P   L   G   L   D   F   H   V   G   L   A   D   E   200

 601 GAG TTT TAT CGC GTG GCT CAT ATA GCG CGC AGT AAA GGC AAT ATG GGC GAT GAA GCC GCA   660
 201 E   F   Y   R   V   A   H   I   A   R   S   K   G   N   M   G   D   E   A   A   220

 661 CAA CGT TTA CTG CAA GTA ATG ATG CGT GAA CCG AAC GCC ATG ACG ACG CGG GCA TTT GCC   720
 221 Q   R   L   L   Q   V   M   M   R   E   P   N   A   M   T   T   R   A   F   A   240

 721 AAC CCA CCT TCT ATT CTG ACC AGC ACC AAT AAA CCC GAA TGG CGA CGC ATG CAG CAG CCG   780
 241 N   P   P   S   I   L   T   S   T   N   K   P   E   W   R   R   M   Q   Q   P   260

 781 GCG GCA AAT GGT CAC GGT AAT GCG CGC AGC CTG GCG GGT TTT TAT AGC GGT TTG TTG GAC   840
 261 A   A   N   G   H   G   N   A   R   S   L   A   G   F   Y   S   G   L   L   D   280

 841 GGT AGT TTG CTG GAA AGC GAC ATG CTC GAA CAA TTG ACT CGT GAA CAC AGT ATT GGG CCG   900
 281 G   S   L   L   E   S   D   M   L   E   Q   L   T   R   E   H   S   I   G   P   300

 901 GAT AAA ACC TTA TTG ACT CAG ACC CGT TTT GGC CTG GGC TGC ATG TTG GAT CAA CCG CAG   960
 301 D   K   T   L   L   T   Q   T   R   F   G   L   G   C   M   L   D   Q   P   Q   320

 961 TTG CCC AAT GCA ACC TTC GGC CTT GGC CCG CGT GCG TTT GGG CAC CCT GGC GCG GGT GGT  1020
 321 L   P   N   A   T   F   G   L   G   P   R   A   F   G   H   P   G   A   G   G   340

1021 TCG GTG GGG TTT GCC GAC CCT GAA CAT GAT GTT GCG TTC GGT TTT GTG ACT AAT ACG TTG  1080
 341 S   V   G   F   A   D   P   E   H   D   V   A   F   G   F   V   T   N   T   L   360

1081 GGG CCG TAT GTA CTT ATG GAC CCT CGT GCG CAG AAG TTG GTC GGA ATA TTG GCC GGT TGT  1140
 361 G   P   Y   V   L   M   D   P   R   A   Q   K   L   V   G   I   L   A   G   C   380

1141 CTG TAA                                                                        1146
 381 L   *
```

Fig. 4. Nucleotide and deduced amino acid sequence of the esterase from *P.fluorescens* KCTC1767. The consensus sequences conserved in the related enzymes are underlined.

TAA. The calculated molecular mass was in good agreement with those determined by SDS–PAGE and gel filtration. A relatively high G + C content (65%) of the ORF also supported its origin being from a strain of *Pseudomonas* (Stover *et al*., 2000). As expected, the sequence showed a distinct amino acid homology (82–93%) and similar structural organization to the mined enzymes through the entire region, as revealed in the identical or counterpart enzymes. The consensus sequences

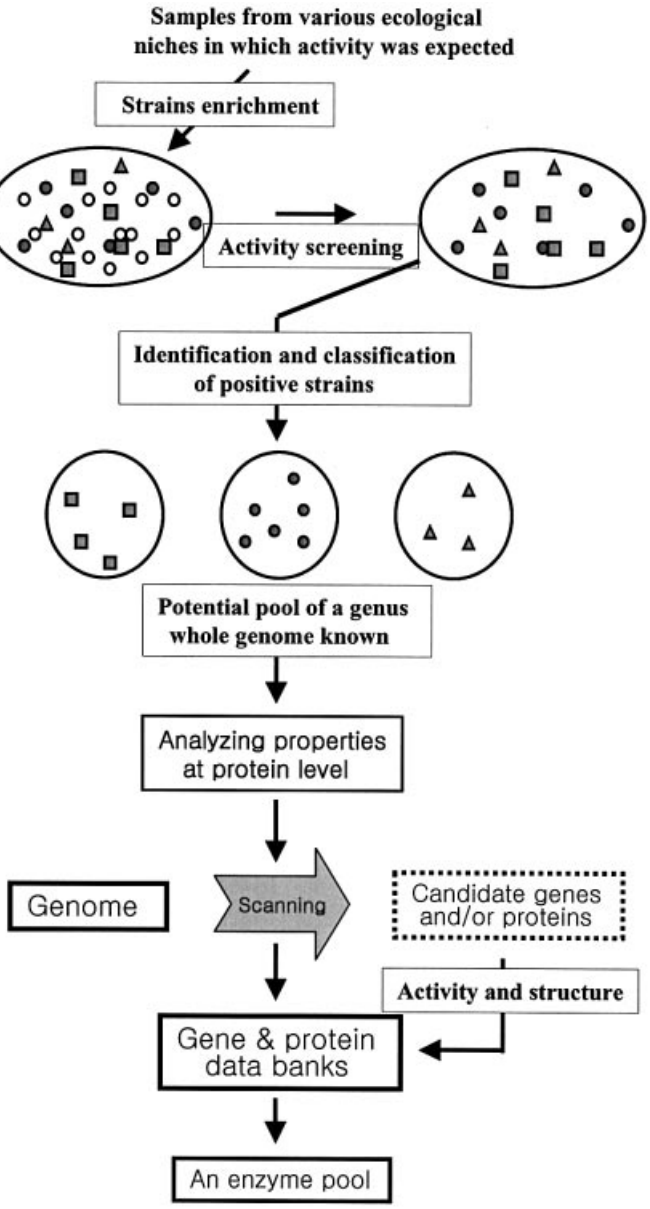

Fig. 5. Typical procedure of the combined approach for the selection of a potential enzyme pool. In this procedure, diverse strain pools were first prepared by cell enrichment and activity screening. After identification, a potential strain pool was chosen and then analyzed for the properties of suspected enzymes. The information gathered will be a basis for the mining of probable gene(s) or enzymes from huge data banks.

conserved in the majority of esterases, H–G and G–X–S–X–G, were also found at the identical positions. These results also strongly suggest that the systematically mined esterases might be able to hydrolyze *rac*-ketoprofen ethyl ester to (*R*)- or (*S*)-ketoprofen, although this remains to be proven. These results will assist us in exploiting the enzymes to obtain further information about their structure and function (Henke and Bornscheuer, 1999).

We attempted here a systematic approach to screen and then express an enzyme with high potential by employing a combined strategy and evaluating it using *rac*-ketoprofen ethyl ester as a target chemical. In brief, the approach consisted of the consecutive steps outlined in Figure 5. Basically, the classical, but still important, processes for the typical screening of enzymes from natural sources are applied prior to analyzing further the suspected pools for preliminary characterization. The subsequent procedures rely on the structural and functional characterization of the suspected pool of enzymes, which can help us to set the criteria for gene mining. Either minor or major activity from a potential pool of strains should be the target for screening and expression. After the characterization of the suspected enzyme pools, numerous programs and search tools for gene and genome data banks are well matched with our intention of mining a valuable gene or enzyme from annotated sequences (Thornton *et al*., 2000; Todd *et al*., 2001). This is currently an easy step because the genome projects of more than 60 strains have been completed to date and many draft sequences are also available in the projects in progress (Clayton *et al*., 1998; Haft *et al*., 2001). Thus, if a strain is chosen as a possible source for an activity, a typical genome will be a basis for gene mining when the conservative evolutions in the related sequences within the related genus are considered. This step is more readily accomplished when the tracing enzyme is a typical one and well known (Koonin and Galperin, 1997). As mentioned earlier, it is still difficult to screen a potential enzyme by typical activity screening and selection (Seed, 1995; Dalboge and Lange, 1998; Ogawa and Shimizu, 1999) owing to the tight regulation and repression of activity under normal physiological conditions, whether the activity is induced or not. Therefore, gene mining and then expression cloning by PCR may provide many advantages over conventional whole cell-based screening.

Currently, the screening or selection of potential enzymes as final products of related research requires the increasing collaboration of skillful experts and thus various fields of activities and tools are being combined, although this has not been fully established. Therefore, a systematic approach combining pre-existing tools is more effective in the screening of an enzyme from natural pools and it is constantly necessary to use a well organized scheme or strategies for concerted effects. An important aspect of this work is the fact that the keroprofen ethyl ester-hydrolyzing genes, including their encoded proteins, are first reported, although the possibility that the reaction may be mediated by an esterase has been suggested elsewhere. However, there is no report on the responsible genes to date. The finding that related compounds, including *rac*-ketoprofen ethyl ester, can be hydrolyzed in a stereospecific manner by the mined esterase enzymes will be a basis for further screening of related enzymes or genes and also will be useful for protein engineering, especially family-driven directed evolution. These approaches may be easily realized from our finding that the genes analyzed here are ubiquitously found in nature.

Given the need to hydrolyze stereospecifically an unnatural substrate, such as *rac*-ketoprofen ethyl ester, that could not induce activity and also could not be used easily as a indicative substrate, the principal question is what strategy is likely to be most effective. In instances in which the esterase-producing cell does not grow well in culture with such a substrate and does not correlate well between the potential and activity of an enzyme, alternative approaches must be used. The enzyme selected here may not be found if the screening is conducted by an activity-based approach, especially when using ketoprofen ethyl ester for the primary screening. As a promising result, we anticipate that it is indeed possible, by using the mined enzymes, to produce an economically valuable drug, (*S*)-

ketoprofen, and also equally valuable products, ibuprofen and naproxen.

Acknowledgements

This work was supported by the Korea Research Foundation Grant (KRF-2001-041-E00352) and the Ministry of Education through the BK21 program. We acknowledge financial support (to G.-J.K) by the 21C Frontier Microbial Genomics and Applications Center Program, Ministry of Science and Technology (Grant MG02-0301-001-1-0-2).

References

Bornscheuer,U.T. (2002) *FEMS Microbiol. Rev.*, **26**, 73–81.
Bornscheuer,U.T. and Kazlauskas,R.J. (eds) (1999) *Hydrolases in Organic Synthesis – Regio- and Stereoselective Biotransformations*. Wiley-VCH, Weinheim.
Caldwell,J., Hutt,A.J. and Fournel-Gigleux,S. (1988) *Biochem. Pharmacol.*, **37**, 105–114.
Clayton,R.A., White,O. and Fraser,C.M. (1998) *Curr. Opin. Microbiol.*, **5**, 562–566.
Cuff,J.A. and Barton,G.J. (2000) *Proteins*, **40**, 502–511.
Dalboge,H. and Lange,L. (1998) *Trends Biotechnol.*, **16**, 265–272.
Fernandez,L., Beerthuyzen,M.M., Brown,J., Siezen,R.J., Coolbear,T., Holland,R. and Kuipers,O.P. (2000) *Appl. Environ. Microbiol.*, **66**, 1360–1368.
Gold,L., Pribnow,D., Schneider,T., Shinedling,S., Singer,B.S. and Stormo,G. (1981) *Annu. Rev. Microbiol.*, **35**, 365–403.
Haft,D.H., Loftus,B.J., Richardson,D.L., Yang,F., Eisen,J.A., Paulsen,I.T. and White,O. (2001) *Nucleic Acids Res.*, **29**, 41–43.
Hayball,P.J. (1996) *Drugs*, **52**, 47–58.
Henke,E. and Bornscheuer,U.T. (1999) *Biol. Chem.*, **380**, 1029–1033.
Jaeger,K.E., Dijkstra,B.W. and Reetz,M.T. (1999) *Annu. Rev. Microbiol.*, **53**, 315–351.
Kim,G.J., Cheon,Y.H., Park,M.S., Park,H.S. and Kim,H.S. (2001) *Protein Eng.*, **14**, 647–654.
Kim,G.J., Choi,G.S., Kim,J.Y., Lee,J.B., Jo,D.H. and Ryu,YW. (2002) *J. Mol. Catal. B*, **17**, 29–38.
Kim,M.G., Lee,E.G. and Chung,B.H. (2000) *Proc. Biochem.*, **35**, 977–982.
Kim,Y.S., Lee,H.B., Choi,K.D., Park,S. and Yoo,O.J. (1994) *Biosci. Biotechnol. Biochem.*, **58**, 111–116.
Koonin,E.V. and Galperin,M.Y. (1997) *Curr. Opin. Genet. Dev.*, **6**, 757–763.
Mauleon,D., Artigas,R., Garcia,M.L. and Carganico,G. (1996) *Drugs*, **52**, 24–46.
McKay,D.B., Jennings,M.P., Godfrey,E.A., MacRae,I.C., Rogers,P.J. and Beacham,I.R. (1992) *J. Gen. Microbiol.*, **138**, 701–708.
Ogawa,J. and Shimizu,S. (1999) *Trends Biotechnol.*, **17**, 13–21.
Park,H.S. and Kim,H.S. (2000) *J. Bacteriol*., **182**, 573–580.
Patal,R.N. (ed.) (2000) *Stereoselective Biocatalysts: Stereoselective Biocatalysis for Synthesis of Some Chiral Pharmaceutical Intermediates*. Marcel Dekker, New York, pp. 87–90.
Seed,B. (1995) *Curr. Opin. Biotechnol*., **5**, 567–573.
Shen,D., Xu,J.H., Gong,P.F., Wu,H.Y. and Liu,Y.Y. (2001) *Can. J. Microbiol*., **47**, 1101–1106.
Stover,C.K. *et al.* (2000) *Nature*, **406**, 959–964.
Thompson,J.D., Higgins,D.G. and Gibson,T.J. (1994) *Nucleic Acids Res.*, **22**, 4673–4680.
Thornton,J.M., Todd,A.E., Milburn,D., Borkakoti,N. and Orengo,C.A. (2000) *Nat. Struct. Biol.*, **7**, Suppl., 991–994.
Todd,A.E., Orengo,C.A. and Thornton,J.M. (2001) *J. Mol. Biol.*, **307**, 1113–1243.
Yuen,P.S., Brooks,K.M. and Li,Y. (2001) *Nucleic Acids Res.*, **29**, E31.

Received July 4, 2002; revised January 20, 2003; accepted April 4, 2003