Journal of Coastal Development ISSN: 1410-5217
Volume 8, Number 2, February 2005 : 49-74 Accredited: 23a/DIKTI/Kep/2004

Review

NOVEL MOLECULAR METHODS FOR DISCOVERY AND ENGINEERING OF BIOCATALYSTS FROM UNCULTURED MARINE MICROORGANISMS

Agustinus Robert Uria[*] *, Yusro Nuri Fawzya, and Ekowati Chasanah*

Research Center for Marine and Fisheries Product Processing and Biotechnology
Agency for Marine and Fisheries Research, Ministry of Marine Affairs and Fisheries
Jl. K.S. Tubun Petamburan VI, Central Jakarta 10260, Indonesia

Received: December, 24, 2004 ; Accepted: January, 20, 2005

ABSTRACT

Metagenomics is a powerful cultivation-independent approach, which can be applied to gain access to the biocatalysts from uncultured marine microorganisms. Discovery of marine biocatalysts by this approach, in general, involves four main steps. First, a metagenomic library containing a pool of biocatalyst-encoding genes is constructed from a marine environment, which can be done by various methods, including cloning of enzymatically-digested DNA, uncut DNA, and PCR-amplified products. Second, the metagenomic library is screened for the genes of interest by employing the activity assay of expression product, in situ hybridization, or Polymerase Chain Reaction (PCR). Third, the obtained target genes, both functional and phylogenetic genes, are sequenced and analysed by using bioinformatic tools in order to gain information on the functional and structural properties as well as the microbial sources of the encoded biocatalysts. Finally, the target genes are expressed in suitable microbial hosts, thereby producing the corresponding recombinant biocatalysts. All existing methods in engineering of marine biocatalysts for the performance improvement can be classified into two main strategies: (i) rational design and (ii) directed evolution. Rational design, which may include the use of resctriction enzyme(s) and splicing by overlap extension (SOE), requires information on the biocatalyst`s structural and functional properties to alter specific amino acid(s). Whereas directed evolution, including error-prone PCR technique and gene shuffling, needs no such information.

Keywords: Metagenomics, biocatalysts, uncultured marine microorganisms, protein engineering

[*]**Correspondence**: Phone: (021) 53650157, 53650158; Fax: (021) 53650158
E-mail: uria_biotek@yahoo.com

INTRODUCTION

Enzymes, which are also known as biocatalysts, are enjoying increasing popularity in the chemical industry as environmentally friendly, economical, and clean catalysts in various applications, ranging from food-related conversion, laundry detergent and paper processing to the production of useful fine-chemicals and research reagents (Wahler and Reymond, 2001). By employing recombinant DNA technology, their amount can be increased in a relatively short time for large-scale applications. For such reasons, it appears that they are gradually taking over from chemical catalysis in many industrial applications (van der Oost and de Graff, 2001).

Recently, biocatalysts originated from marine microorganisms have attracted much scientific attention for being used in various industrial processes, because they possess

unique properties which are industrially important. For example, commercially valuable enzymes (e.g. proteases, xylanases, amylases, lipases, and esterases) from marine hyperthermophilic microorganisms had been found to be thermally stable and resistant towards denaturing agents (Bertoldo and Antranikian, 2002). Therefore, it seems that marine microorganisms become a rich reservoir of interesting new biocatalysts.

For many years, discovery of new biocatalysts from microorganisms are mostly based on cultivation-dependent methods, in which microorganisms are enriched and isolated in pure cultures, followed by extracting the biocatalysts of interest. However, these methods count only less than 1 % of the microbial diversity in most environments, including marine environment. As a result, almost all the biocatalysts in use today came from the small fraction of microorganisms that can only be grown under laboratory conditions (Arnold, 2001).

A recent breakthrough in the area of Microbial Ecology has led to the birth of a powerful approach for gaining access to the enormous uncultured microbial diversity (Riesenfeld, *at al*. 2004). This cultivation-independent approach, popularly known as metagenomics, provides a faster way of discovering an unlimited pool of new potential biocatalysts, genes, and biosynthetic pathways from uncultivable marine microorganisms (Arnold, 2001; Daniel, 2001). More recently, many single genes, such as those coding for cellulolytic enzymes (Handelsman, 2004) and for chitinases (Cottrell, *et al,* 1999, 2000), have been successfully collected from environments and expressed by using this metagenomics approach.

In spite of the potential of finding new biocatalyst from marine micro-organisms, however when biocatalysts from nature are recruited for industrial applications, it often appears that their performance needs to be improved. This is intended to create enzyme mutants with desired properties, such as stable over long periods of time, active in unnatural and non-aqueous solvents, and able to accept specific substrates of interest (van der Oost and de Graaff, 2002). Improving biocatalyst's performance can be achieved by engineering them at molecular level. In general, all of the existing methods for biocatalyst engineering can be divided into two main groups: site-directed mutagenesis (rational design) and random mutagenesis (laboratory/directed evolution). Both groups will be briefly discussed in this paper.

DISCOVERY OF BIOCATA-LYSTS FROM UNCULTURED MARINE MICROORGANISMS

Search for new biocatalysts from uncultured marine microbes through metagenomics, in principle, involves four following stages (Handelsman, 2004) (**Fig. 1**): (i) construction of metagenomic library, (ii) screening of metagenomic library, (iii) sequence analysis of genes, and (iv) expression of genes of interest in appropriate hosts, thereby producing the corresponding biocatalysts. In the first stage, genes are collected from a marine environmental sample. This is normally done by isolating total DNA from the sample, followed by its partial digestion with a restriction enzyme. The digestion products, DNA fragments, are combined with a specific DNA molecule called vector. The result is called recombinant DNA molecules which are also known as clones. Whereas, the individual DNA fragments incorporated in clones are usually named DNA inserts. Such clones are then introduced into a microbial host, usually *Escherichia coli*, through either heat shock or electric shock. This procedure is widely called transformation; and the products are named transformed cells (transformants), which are collectively known as a metagenomic library. The DNA insert present in a transformant may contain functional gene(s), phylogenetic gene(s), or both of them. Functional gene means the gene coding for a corresponding protein/ biocatalyst. Whereas a phylogenetic marker, usually 16S rRNA gene, is not expressed into protein.

The metagenomic library constructed needs to be screened in order to detect transformants containing functional genes as well as phylogenetic markers. If the

functional genes present in transformants are expressed, such screening can be done by assaying the activity of the expression products, in this term biocatalysts (enzymes). However, in the case that the genes are not expressed, thereby no corresponding enzymes are produced, employing *in situ* hybridization or polymerase reaction chain (PCR) might be the right choice. These methods enables to detect target genes present in each transformant.

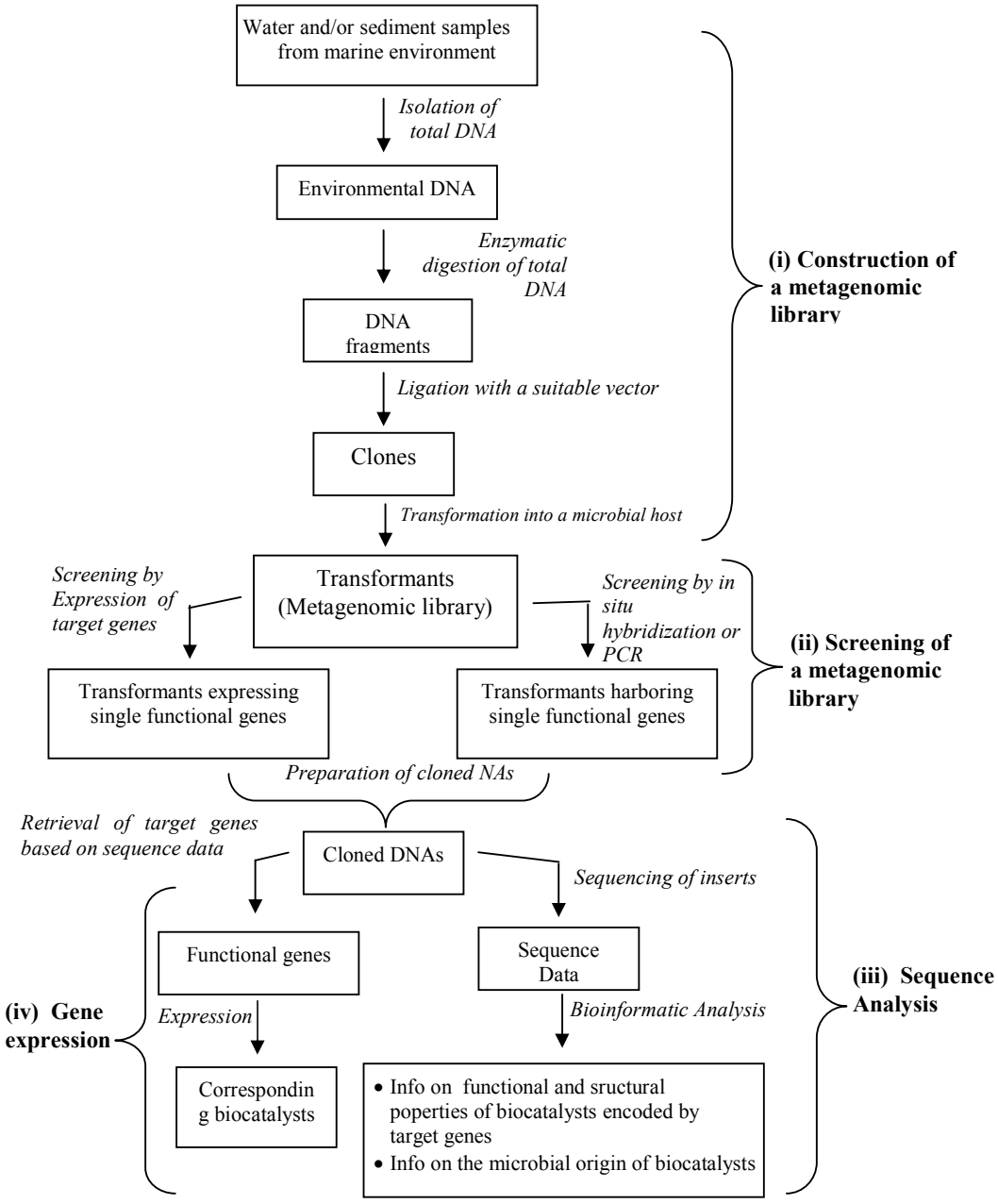

Fig. 1. Scheme of metagenomics for the discovery of new biocatalysts from uncultured marine microorganisms.

After detecting transformants containing foreign genes, the individual inserts harboring genes need to be recovered from the transformants and then sequenced. The sequence data are analyzed by using the existing bioinformatics tools and databases, which are freely available on internet. The analysis results can give a useful information on the functional and structural properties of the biocatalysts encoded by corresponding genes. If the properties are different from those of the known biocatalysts available in the databases, it means that the cloned genes and corresponding biocatalysts are newly found, which are not reported yet previously.

In the case that a functional gene is located in the same single DNA insert as a phylogenetic marker (16S rRNA gene), sequence analysis of the marker can give a clue about the microbial source of the functional gene. The 16S rRNA gene is now widely used as an appropriate marker for bacterial identity, because its transcription products (16S RNAs) are excellent molecules for discerning evolutionary relationship among living organisms. This is based on some chronometric properties that RNAs possess, such as functionally constant, universally distributed, and moderately well conserved in sequence across broad phylogenetic distances. Therefore, similarity in rRNA sequences between two microbial organisms indicates their relative evolutionary relatedness (Madigan, *et al.*, 2000).

So far the number of microbial biocatalysts discovered by the metagenomic approach is still relatively small in comparison to that by the classical cultivation-dependent approach. However, recent studies have shown a large number of marine biocatalyst-encoding genes and phylogenetic markers which were screened from uncultured microorganisms by the metagenomic way, such as lipases (Henne, *et al*., 2000), amylases (Rondon, *et al.*, 2000), chitinases (Cottrel, *et al*., 1999), and esterases (Henne, *et al*., 2000). Some of them originated from uncultured marine microorganisms, as listed in **Table 1**.

Table 1. Some marine biocatalyst-encoding genes and phylogenetic markers discovered through metagenomics.

Marine biocatalyst-encoding gene and phylogenetic marker	Vector	Marine microbial source	Reference
Gene coding for Ribulose-1,5-bis-phosphate carboxylase/ oxygenase (RubisCO), involved in carbon fixation	Fosmid	Deep-sea tube warm symbionts	Handelsman (2004)
The genes *onn*A-*onn*H encoding poly-ketide synthase (PKS) involved in the biosynthesis of onnamide antitumor.	Cosmid	Sponge symbiont	Piel, *et al* (2004)
Methyl Coenzyme M Reductase (MCR) Archael 16S RNA gene	Fosmid	Marine Sediments	Hallam, *et al* (2003)
Histidin protein kinase gene	Fosmid	Tubeworm symbionts	Hughes, *et al* (1997)
Archael 16S rRNA gene		Sponge symbiont	Schleper, *et al* (1998)
ATP citrate lyase 16S rRNA	Fosmid	Deep-sea poly-chaete symbionts	Champbell, *et al* (2003)
16S rRNA	Plasmid	Marine coastal environment	Tanner, *et al* (2000)
Chitinases	Phagemid	Coastal seawater	Cottrell, *et al* (1999)
Chitinases	Plasmid	Coastal and estuarine water	Cottrell, *et al* (2000)

1. Construction of Metagenomic Library

Construction of metagenomic libraries from a marine environment for the discovery of biocatalysts can be performed by various molecular methods. The common method is based on shotgun cloning, which is successfully employed for obtaining a variety of enzymes, such as alcohol dehydrogenase as reported by Knietsch *et al*. (2003). Total DNA was directly retrieved from an environmental sample without prior cultivation, which was then partially digested by using the restriction enzyme, *Bsp*1431. The resulting DNA fragments are joined with the vector, pBluescript SK^+, and the clones formed was tranformed into *E. coli* DH5α for clone amplification, which was followed by transfromation into *E. coli* ECL707 for insert expression. The expression product was prepared from the bacteria at the stationer phase and assayed for alcohol dehydrogenase activity.

An impressive breakthrough in construction of metagenomics libraries is the success in the recovery of high-molecular-weight DNA from soils as reported by Bertrand *et al*. (2005). This offers a big chance for constructing a relatively large library through shotgun cloning, as reported by Rondon *et al.* (2000). Two large metagenomic libraries containing more than 1 Gbp of DNA insert was constructed by using bacterial artificial chromosome (BAC) vector. Subsequently, genes coding for 16S rRNA were retrieved from the libraries using PCR and then sequenced. The resulting sequence was compared with the sequence database at the National Center for Biotechnology Information (NCBI) using BLAST software. Then the libraries were secreened to find some commercial biocatalysts. The comparison results showed that the DNA inserts (environmental DNA) originated from a wide diversity of microbial phyla, including gram-positive *Acidobacterium, Cytophagales,* and *Proteobacteria.* The screening results revealed the presence of genes encoding lipase, amylase, and nuclease.

Another interesting method for constructing a large metagenomic library was reported by Courtois *et al.* (2003). This method involved the direct isolation of high-molecular-weight DNA (±50 kb) from the environment. The DNA sample obtained was not digested with any restriction enzyme, but added with polynucleotide tails (poly-dT) in order to increase the change of gaining the gene clusters encoding the multienzyme, polyketide synthase (PKS). Subsequently, the tailed and uncut DNA was joined with a cosmid vector. Before such joining, at first the vector was linearized by using a restriction enzyme and tailed with poly-dA by the use of the enzyme, terminal deoxynucleotidyl transferase. The resulting recombinant DNA molecules were transferred into *E. coli* cells through λ phage infection, thereby resulting in a relatively large metagenomic library. The desired recombinant colonies on the library, marked with white color on agar plates, could be picked and gridded to be compatible with 96-well microtiter plates containing freezing medium (Luria-Bertani, 20% glycerol complemented with 12.5 μg of chloramphenicol/ml), as suggested by Rondon, *et al.* (2000) and Ginolhac, *et al.*(2004). After growing at 37^{o}C for 22 hrs with 250 rpm shaking, the plates are stored at -20^{o}C

Another method is based on the use of PCR with specific primers to isolate and amplify a gene of interest from environmental genomic DNA. The PCR products are cloned into a bacterial host. This method had been reported by Cottrell *et al*. (2000) for chitinase genes of marine microbial origin. Seawater samples were sieved by using 0.22-μm filtration catridge to obtain bacterial cells without prior cultivation. Then, DNA was isolated from such cells; and chitinase genes were isolated and amplified from the DNA by using PCR. The degenerate primers used for the PCR were the forward primer IICRFORB (5`-wsigtiggiggitggcanyt-3`) and the reverse primer GRP1571AR (3'-ctrcgictrttrccictrta-5`), which were designed based on deduced amino acid sequences of chitinases in four γ-proteobacteria (*Alteromonas* sp., *Aeromonas caviae*, *Serratia marcescens*, *Enterobacter agglomerans*). The resulting PCR products can then be cloned in appropriate bacterial host, thereby generating a metagenomic library.

In the case of building a metagenomic library from the eukaryotic microbial DNA, isolation of total RNA from a marine environmental sample is recommended instead of DNA isolation. This is related to the presence of intervening non-coding sequences (introns) in most of eukaryotic genes, which are not present in the corresponding mature *messenger* RNAs (mRNAs). Then mRNAs are purified from the total RNA and used as the template for synthesizing *complementary* DNAs (cDNAs). The resulting cDNAs are intron-free eukaryotic genes, which can be used further for constructing a metagenomic library. The metagenomic libraries constructed from environmental RNA needs to be selected or screened to identify single genes coding for biocatalysts of interest.

2. Screening of Metagenomic Library

Searching for new biocatalysts in large metagenomic libraries is a challenging stage in metagenomics. This requires high-throughput screening methods which are relatively practical, inexpensive, and fast. In principle, there are some different methods in library screening to detect transformants harboring the biocatalyst-encoding genes of interest. The methods, in principle, are based on (i) detection of the expression product of target gene and (ii) detection of target gene in an insert by employing PCR and hybridization techniques (Handelsman, 2004; Riesenfeld, *et al*. 2004).

The success of the screening method based on detection of the expression product really depends on the ability of transformants to express genes of interest. Therefore, a suitable host is required for this method. There is a variety of strategies to identify desired expression products (biocatalysts) in a high-throughput format. One of them is screening based on the assay of individual transformant colonies present in a library for the ability to express desired biocatalysts (Wahler and Reymond, 2001a). The most simple one is the use of indicator media containing certain substrates to detect the presence of enzymes of hydrolase group. The presence of a hydrolytic enzyme is confirmed with a clearing zone around the positive transformants. For detecting protease, keratinase, chitinase, and lipase activity, the metagenomic library is replicated to Luria-Bertani plates containing the following substrates: nonfat dry milk, keratin powder, chitin powder, or bacto lipid, respectively. The enzyme activities are confirmed by the presence of clearing zones (Rondon, *et al*., 2000).

The most convenient strategies to identify biocatalysts in high-throughput are those involving the use of chromogenic or fluorogenic substrates. These strategies are based on the ability of bacterial transformants to take the substrates selectively and convert them into products which can visually be detected *in vivo*. For example, detection of the transformants expressing *cytochrome* P450 can be performed by using naptalene derivative as the fluorogenic susbtrate (**Fig. 2A**). The hydroxilative product of this enzyme on such substrate is fenol, which is polimerized further by *horse radish peroxidase* (HRP) to form a fluorocent dimer/polymer. Similar strategy can also be employed to monitor the activity of *transglycosidase* on the substrate α-D-glucosylfluoride and *para*-nitrophenyl-β-D-glucoside (**Fig. 2B**). This strategy involves endocelullase to hydrolize the product of *transglycosidase* activity, generating yellow nitrophenol as the indicator for the presence of transglycosidase (Wahler & Reymond, 2001b).

ISSN: 1410-5217
Accredited: 23a/DIKTI/Kep/2004

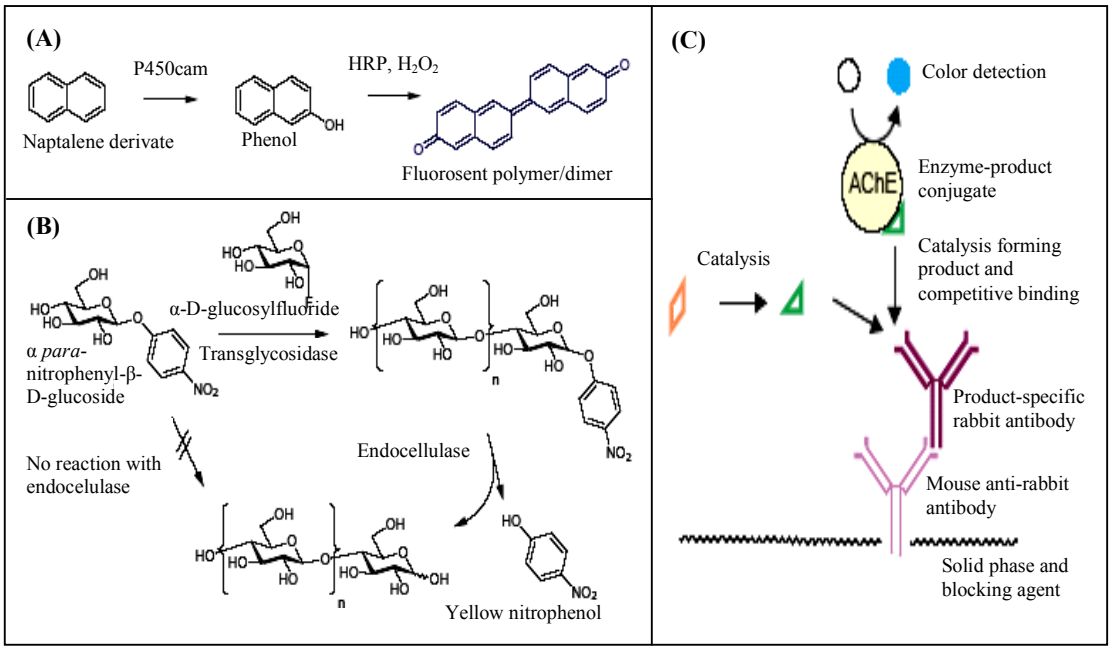

Fig. 2. Some examples of screening strategies for certain biocatalysts. (A) Fluorogenic substrate-based detection of cytochrome P450 activity. (B) Detection of transglycosidase activity by using fluorogenic substrates. (C) Competitive cat-ELISA-based screening for a biocatalyst. Copied from Wahler & Reymond (2001 a and b).

Another interesting screening for biocatalysts is the competitive cat-ELISA assay (**Fig. 2C**) relying on the binding of antibody to test not only for the activity of a biocatalysts, but also its selectivity. This assay is based on the competitive binding between product and enzyme-product conjugate for the occupation of the binding sites of the product-specific antibody (Wahler and Reymond, 2001a). High-throughput screening can also be aimed at the biocatalyst's enantioselectivity as well as its activity. For instance, enzyme assay is based on secondary processes of fluorosence release, as proposed by Reymond (2001). The fluorescent product, umbelliferone is released by *β*-elimination of an intermediate carbonyl compound. This screening type involves substrates which can be converted into a detectable product after a secondary reaction sequence. This assay provides stereo- and enantio-selectivity for alcohol dehydrogenase and aldolases. It also operates for hydrolytic enzymes such as esterases, lipases, amidases, epoxide hydrolases and phosphatases.

In the case that genes in clones are not expressed in host, *in situ* hybridization or PCR can be used. For example, hybridization detection of the multienzymes polyketide synthase in a metagenomic library constructed from soil DNA, as reported by Ginolhac *et al.* (2004). At first, they designed ketosynthase-based primers for isolating specific fragments from soil DNA. Then the resulting fragments were used as probes labeled with [α-^{33}P]dCTP. Subsequently, transformants were transferred on to nylon membranes and lysed to expose the DNA. The DNA on membranes was immobilized; and hybridization with probe was performed. The membranes were washed sequentially; and the hybridization signals were visualized. The position of the signals indicated positive transformants on a master plate.

A recent imprensive breakthrough in the use of PCR technique for screening of

metagenomic library was reported by Piel, *et al.* (2004). They employed a PCR-based screening to obtain the uncultivated bacterial gene cluster encoding a polyketide synthase (PKS) responsible for manufacturing the anti-tumor polyketide, onnamide. The library was constructed from the metagenomic DNA of the uncultivated bacterial symbionts of the marine sponge, *Theonella swinhoei*. For such PCR-based screening, two PCR primers, in this respect sponge3f (5`-tggagctcagttggcaggta-3`) and sponge3r (5'-tggtttcaacagcagcatcac-3`), were designed based on the comparison among the keton synthase (KS) sequences of PKS genes. Subsequent transfer of the complete onnamide-specific gene cluster into easily culturable microbes could produce the marine drug in any desired amount, thereby creating the renewable and bulk supply of rare sponge-derived pharmaceuticals.

3. Sequence Analysis of Biocatalysts

After finding target biocatalysts via screening of metagenomic libraries, the biocatalysts can be characterized based on their amino acid sequences via bioinformatic approach. One factor allowing to do this approach is the exponential growth of public databases containing protein structures and sequences, such as National Center for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov/), European Molecular Biology Laboratory (EMBL, www.EBI.org), Protein Data Bank (PDB, www.pdb.org), and SwissProt (www.expacy.org). This is ussually initiated with the retrieval of DNA fragments (inserts) coding for the biocatalysts from positive clones, which is followed with complete sequencing of such fragments. The resulting nucleotide sequence data can be translated first into the corresponding amino acid sequences. Then by using some bioinformatic softwares, which are freely available in public databases, the amino acid sequences can be analyzed in order to predict a number of biocatalyst's properties (e.g. molecular weight, isoelectric point, etc.) as well as to identify the availability of phylogenetic markers (e.g. 16S rRNA gene).

The sequences coding for biocatalyst of interest can be subjected to *BLAST* (Basic Local Alignment Search Tool) program to search for homologous sequences in databases. Further search of the additional information on the sequences can give a clue about the structural and functional features of the target biocatalysts. If the query sequences are aligned together with the well-known significant homologous sequences through the use of Clustal W or X program, the results can give useful information, such as substrate- and cofactor-binding sites, thermostability, and enantioselectvity (Uria, *et al.*, 2004).

An example of bioinformatic analysis is the prediction of cofactor- and substrate-binding sites on thermostable alcohol dehydrogenase (ADH D) of marine archaebacterial origin, as reported by Uria (2004) and Machielsen, *et. al.* (2006). The primary sequence of the gene coding for ADH D (*adh* D) was retrieved from the database NCBI and then converted into the amino acid sequence using the software, *EditSeq* of DNAStar. The amino acid sequence obtained was subjected to the pairwise alignment (BLAST search) to find its similarity with other sequences in the databases. The sequences showing the significant similarity were aligned along with the query sequence using CLUSTAL W. The results, as shown in **Fig. 3**, revealed that among the conserved residues, five of them seemed to be likely the residues in the majority of aldoketo reductase cofactor-binding sites (Asp_{58}, Asn_{153}, Ser_{267}, Arg_{271}) as marked with open boxes in Fig. 3. Then three of them were identified as substrate-binding sites (Asp_{58}, Tyr_{63}, Lys_{89}, His_{121}) marked with arrows in **Fig. 3**. It was assumed that Tyr_{63} in ADH D seems to function as the catalytic acid, and His_{121} is responsible for orienting the carbonyl of the substrate in the active site. The catalytic reaction is initiated by transferring a hydride ion (H^+) from NAD(P)H to the substrate by the help of ADH D enzyme, followed by proton transfer to the carbonium-ion intermediate.

One of challenges with metagenomics is to identify the original microbial source of the target biocatalysts obtained. This can be achieved by finding a phylogenetic marker

(e.g. 16S rRNA gene) flanking the functional gene of interest. Once a functional gene is linked to phylogenetic markers (e.g. 16S rRNA genes) on a single DNA fragment (insert), it is possible to unravel the microbial source of the gene by comparing the marker sequence with the known 16S rRNA gene sequences in the databases. The use of phylogenetic marker to identify the environmental genomic fragments was initially proposed by the DeLong research group which produced the first genomic sequence linked to a 16S rRNA gene of an uncultured archaebacteria (Handelsman, 2004).

Identification of the fragment of interest based on 16S rRNA gene analysis is ussually initiated by comparing the 16S rRNA gene sequence of a target fragment with known related sequence in GenBank by using advance gapped BLAST software (Tanner, *et al.*, 2000). By performing a multiple alignment between the target sequence and its relate sequences, a phylogenetic tree indicating the microbial origin of the fragment of interest can be constructed. A phylogenetic tree built can be calculated with the neighbour-joining algorithm using the program NEIGHBOR of the PHYLIP version software (Suzuki, *et al.* 1997).

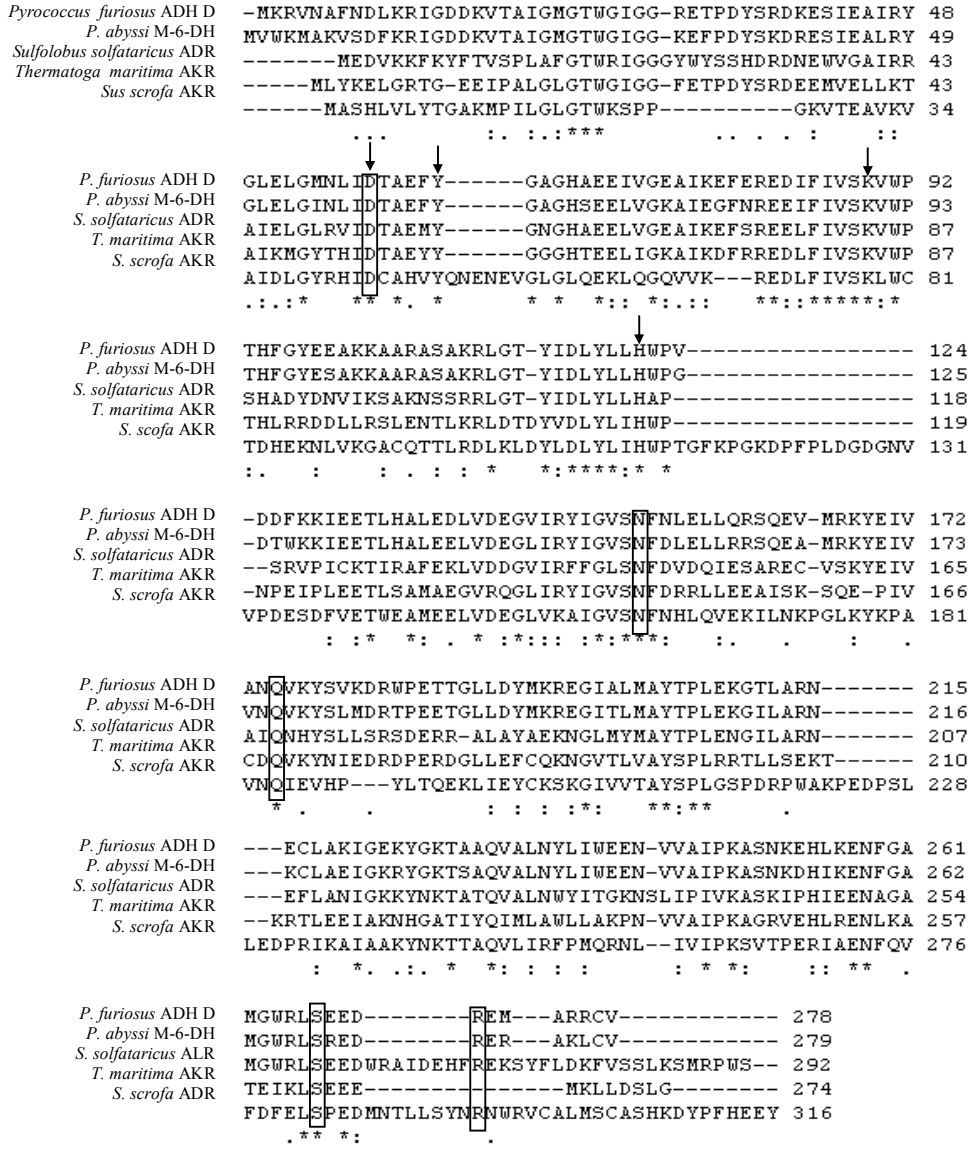

Fig. 3. Sequence comparison between *P. furiosus* ADH D and four enzymes belonging to the aldo-keto superfamily by using the software, Clustal W. ADH D, NAD(P)-dependent alcohol dehydrogenase; M-6-DH, morphine-6-dehydrogenase (80%); ADR, aldehyde reductase (53%); AKR, aldo-keto reductase (69%); ADR, Aldose reductase (48%). Cofactor- and substrate-binding sites were marked with open boxes and arrows (Uria, 2004; Machielsen, *et al.,* 2006)

4. Expression of the Genes of Interest

After target genes of uncultured microbial origin are discovered, the next important step is the expression of such genes in easily cultivated microbial systems, thereby producing the enzymes encoded by the genes. For such purpose, there are several things necessary to be considered such as suitable hosts, vectors, and pro-moters. Promoter is the region on DNA in front of a gene where the RNA polymerase binds to before beginning transcription.

Among the microbial systems available for gene expression, the gram-negative bacterium, *E. coli* is still one of the most attractive and widely used host due to its ability to grow rapidly at high density on inexpensive substrates, its well-characterized genetics, the availability of an appropriate large number of cloning vectors, and the availability of increasingly mutant host strains (Baneyx, 1999). The *E. coli* strains used as the host normally lack restriction enzymes and are *recA*$^-$ to reduce the recombination between the cloned DNA and the host chromosom (Prescott *et al.*, 2002). To express a foreign gene in the *E. coli* host, at first the gene needs to be inserted into a suitable cloning vector, ussually under the control of a promoter. The product is called recombinant DNA (rDNA) molecule, which is further transformed into *E. coli* host. In the host, when the promoter controlling the foreign gene is recognized by the host RNA polymerase, the gene is automatically trancribed into mRNA molecules. The resulting mRNAs can then be translated into enzyme molecules in ribosomes. The enzyme produced by this way is widely called recombinant enzyme.

Among a variety of *E. coli* expression systems commercially available, the pET system (commercialized by Novagen), widely known as one of the most premier and powerful systems for the expression of recombinant proteins, has gained increasing popularity in recent years. This system provides six possible vector-host combinations which enable tuning of the expression levels to optimize target gene expression. One of the widely used vector-host combinations is pET24d coupled with the host, *E. coli* BL21 (DE3) strain (Novagen 2002-2003 Catalog, www.novagen.com). This combination was succesfully used to express an archaebacterial alcohol dehydrogenase-encoding gene (*adh*) in *E. coli* BLD(DE3) strain, as reported by Uria (2004) and Machielsen, *et al.* (2006). Such *adh* expression was done by two following steps: (i) inserting *adh* gene into pET24d vector, thereby generating a recombinant vector designed as pWUR85 (**Fig. 4a**) and (ii) transforming pWUR85 into *E. coli* BLD21(DE3) via electric shock. The desired transformants containing harboring a target gene were screened by taking advantage of the antibiotic-resistance marker and *lac* I gene present in the vector. By supplementing the growth medium with a certain antibiotic, vector-free cells are killed, leaving the transformed cells. Then the presence of *β*-galactose in the agar plates enables to detect and isolate the desired transformants harboring a target gene, which are marked with white colonies. In the desired transformant, the *adh* gene was expressed into the recombinant alcohol dehydrogenase, designated as ADH D, through a molecular mechanism described in **Fig. 4b**. The T7 RNA polymerase gene was under the control of *lac* promoter located on the host chromosome. Binding the *lac* repressor molecules at the *lac* operator effectively reduces the transcription of the T7 RNA polymerase gene, leading to the inhibition of target gene expression. However, when the inducer, isopropyl-β-D-1-thiogalacto-pyranoside (IPTG) was supplemented into the bacterial growth medium, it inactivated the repressor binding, leading to the expression of the target gene, *adh* gene.

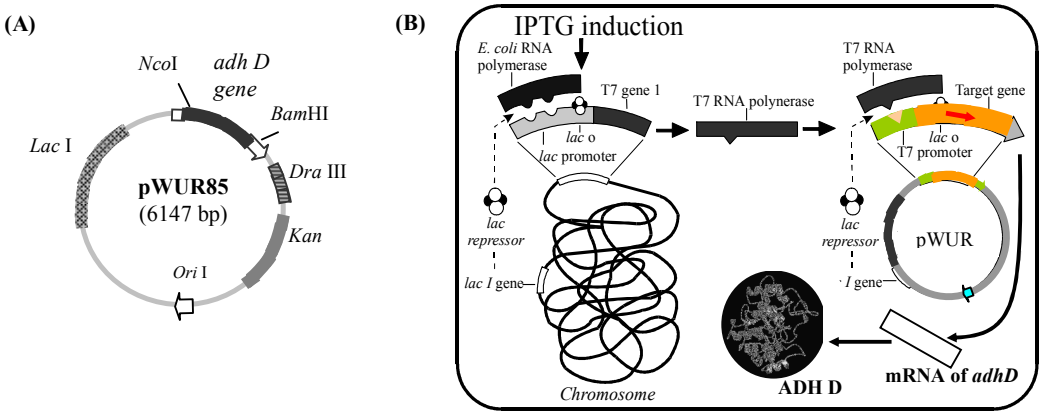

Fig. 4. A vector-host combination of the pET system used for expressing an alcohol dehydrogenase (ADH). **(A)** The expression vector, pWUR85 harbouring *adhD* gene; *Kan*, kanamycine resistance gene; *adh D,* alcohol dehydrogenase D gene; *Lac* I, lactose gene; *Ori* I, origin of replication I. **(B)** Expression of ADH gene in *E. coli* BLD21 (DE3) harbouring the tRNA helper plasmid, pSJS1244 after being induced with IPTG (Uria, 2004).

At present, the exploitation of hyperthermophilic archaebacteria as cell factories for overexpressing interesting heterologous biocatalysts is still in the early stages of development. So far, among the extensively studied hyperthermophiles, *Sulfolobus, Pyrococcus* and *Thermococcus* are good candidates as the production hosts. This is because of the availability of their genome sequences in the public databases, enabling the full development of metabolic engineering. Another reason that they are relatively easy to handle and can be grown to high cell density (Lucas, *et al.,* 2002).

Recently, it has been reported by Contursi, *et al.* (2003) about the successful expression of thermostable ADH in *S. solfataricus* by using the vector, pEXSs. In addition, a pyroccocal shuttle vector designated as pYS3 (**Fig. 5**) was successfully established by Lucas, *et al.* (2002). The vector was equipped with replication elements and genetic markers which allow it to work well both in the *E. coli* and pyrococcal systems. The replication elements in the vector are CoIE1 origin of replication (*Ori*) for the *E. coli* host and the *rep75* for the pyrococcal host. Whereas, the genetic markers are ampicillin resistant-coferring gene (*Ap* gene) for the *E. coli* host and orotate phosphoribosyltransferase gene (*pyrE*) and orotidine 5-monophosphate decarboxylase gene (*pyrF*) for the pyrococcal host (Lucas, *et al.*, 2002).

In another hand, strong promoters as the critical elements for gene expression have been isolated from various microorganisms. One of very promising future promoters is the *Psip* promoter, in which its function is strongly controlled by elemental sulfur (Schut, *et al.*, 2003). This promoter have succesfully been isolated and amplified from *P. furiosus* genome by using PCR and then combined with a pyrococcal *adh* gene, as reported by Uria (2004). For *adh* expression in an archaebacterial system, this *adh-Psip* hybrid needs to be incorporated into the vector pYS3; and the resulting recombinant vector can then be transferred into the *E. coli* BL21(DE3) for amplifying and selecting the *adh* gene. As described in Fig. 5, the *E. coli* transformants containing such recombinant vector can be screened by a negative selection on agar plates containing-fluoroorotic acid (5-FOA). Such

ISSN: 1410-5217
Accredited: 23a/DIKTI/Kep/2004

transformants are unable to grow in the presence of 5-FOA; whereas the non-transformants become resistant to 5-FOA due to the deficient in one of these *pyr* genes. Based on the characteristics, the transformants can be detected on the master plates (Lucas, *et al.,* 2002). The recombinant vectors containing such gene are then recovered from the transformants and cloned into the hyperthermophilic host, *P. abysii*. By supplementing S^o into the growth medium, the promoter is activated, leading to the expression of the target gene.

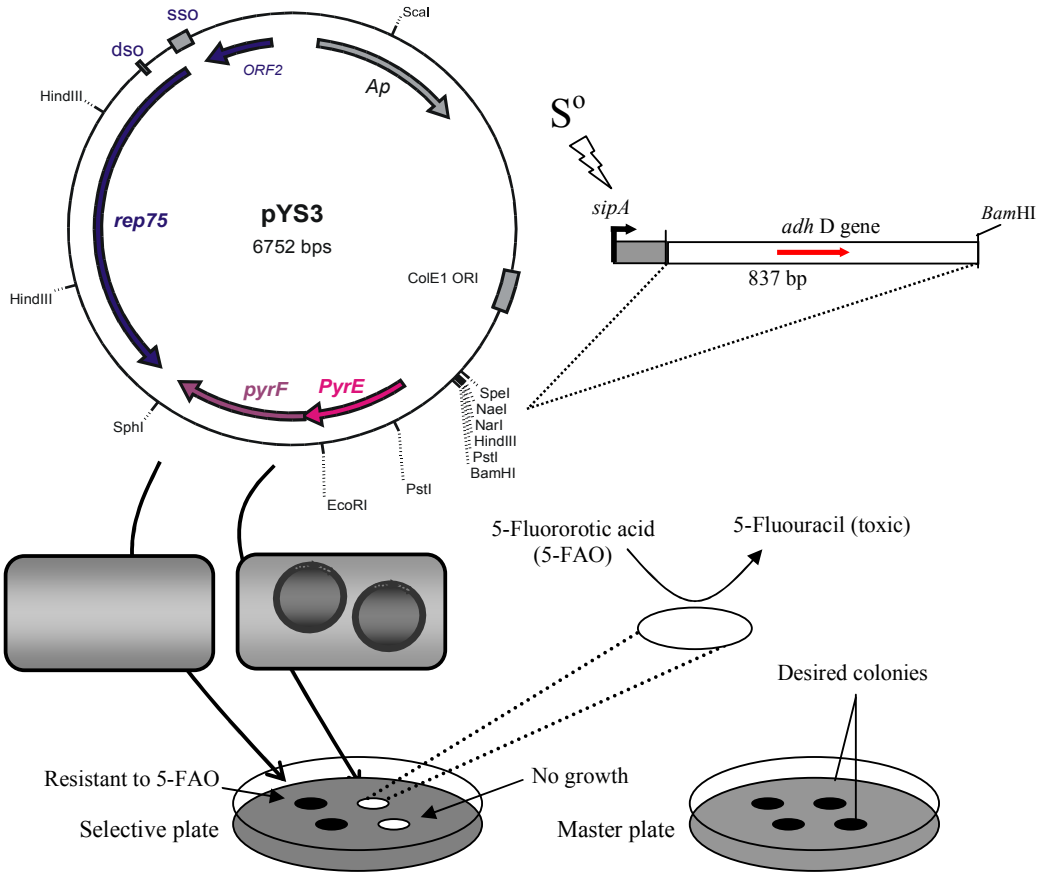

Fig. 5. A negative selection-based strategy for detecting the desired transformants harboring a recombinant shuttle vector

MOLECULAR ENGINEERING OF BIOCATALYSTS

Genetic engineering of biocatalysts can be performed via two different approaches: (i) rational protein design and (ii) laboratory evolution. For employing both approaches, there are several important components required, including the single genes of interest, a microbial expression system, and a sensitive detection/screening system. Both approaches will be discussed in this section.

1. Rational Design

Rational design is a still very effective site-directed mutagenesis strategy to elaborate

improved proteins. This strategy needs the information on the structure of protein or enzyme as well as the relationship between sequence, structure and function (Bornscheuer and Pohl, 2001). As mentioned previously, when the amino acid sequence of a target biocatalyst has been determined, the sequence can be processed to search for similar sequences from public databases. A multiple alignment between the target sequence and the functionally-known homologous sequences can give useful information on residues responsible for biocatalyst functional properties such as substrate- and cofactor-binding sites, thermostability, and enantioselectivity. This may become a basis for performing rational design on target proteins.

There are various existing methods in rational design to improve biocatalyst performance. One of them is making a specific alteration on the critical residues by using restriction enzymes. This method had been successfully demonstrated in improving the stability of a *Bacillus* serin protease towards chemical oxidation, as reported by Estell, *et al.* (1985). In their experiment, the restriction enzymes, *Pst*I and *Kpn*I, was employed to remove the oxidation-sensitive methionine 222 and replace it with nonoxidizable amino acids (*i.e.* Ser, Ala, and Leu). The protease mutants containing such amino acids were found to be resistant to chemical oxidation. Another example indicating the success of using this method is enhancing the thermostability of α-amylase and *D*-xylose isomerase by replacing asparagines residues with the more rigid structural residues, such as proline (Bornscheuer and Pohl, 2001). A similar result was also reported by Van den Burg, *et al*. (1998) for the moderately stable enzyme, thermolysin-like protease from *Bacilus stearothermophilus*.

Among various methods in rational protein design, splicing by overlap extension (SOE) is a relatively simple and rapid PCR-based method for generating improved protein variants containing a specific amino acid alteration (**Fig. 6**). This method offers significant advantages, including not using restriction enzymes or DNA ligase, the ability to introduce site-specific mutations, and the extreme ease, speed, and reproducibility in generating variants (van der Oost and de Graaff, 2002). This method, in principle, involves designing two overlapping mutagenic oligonucleotides as PCR primers, and application of PCR and both primers on a wild-type gene, leading to the generation of gene variants with a specific mutation. The resulting gene variants are joined with a suitable expression vector and transformed into a bacterial host. When the gene variants are expressed in the host, a large quantities of mutated biocatalyst is produced.

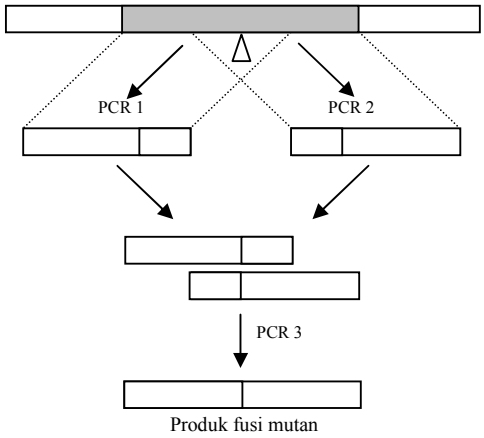

Fig. 6. Site-directed mutagenesis by SOE which involves PCR-amplification of a wild-type gene using two overlapping mutagenic primers. The PCR products of PCR reaction 1 and 2 are overlapping fragments, which are joined together in the third PCR reaction, thereby generating mutated variants of wild-type protein. Copied from van der Oost and de Graff, 2002.

2. Laboratory Evolution

Recently, random mutagenesis by laboratory evolution, which is also called directed evolution or molecular evolution, has emerged in just a few years as one of the most effective approaches to develop and tailor biocatalysts for industrial requirements. Laboratory evolution, in principle, relies on a laboratory-scale imitation of Darwin evolution. By this approach, a series of biocatalyst properties have been improved, in the absence of detailed structural and functional information, including stability, catalytic activity, and specificity towards new substrates (Arnold, 2001).

The first step in laboratory evolution is creating an immense diversity of molecular variants from a wild-type gene by introducing point mutation and/or by recombination. The resulting molecular variants are screened to find improved variants with desired properties. Introducing point mutation can be performed by using error-prone PCR technique (**Fig. 7**), in which divalent cations (e.g. Mn^{2+}) is added in PCR reaction to substitute the optimal Mg^{2+} or high error-rate polymerase (e.g. *Taq* polymerase lacking the ability of proofreading) is employed (van der Oost and de Graff, 2002; Arnold, *et al.,* 2001).

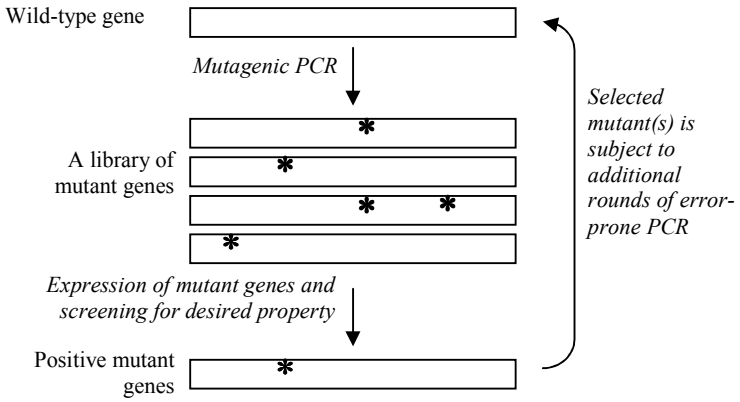

Fig. 7. Laboratory evolution by error-prone PCR procedure, which involves PCR-based creating novel random mutants from a wild-type gene. After expressing the individual mutants in a bacterial host, the resulting transformants are screened for mutated biocatalysts with a desire property. To improve the desire property, the biocatalysts can be subjected to additional rounds of error-prone PCR (van der Oost and de Graff, 2002).

Another attractive method in performing laboratory evolution is DNA shuffling, which is useful to generate highly improved proteins/biocatalysts with desired features. This method, which was initially implemented by Stemmer (1994), involves the fragmentation of a pool of homologous genes and the PCR-reassembly of the resulting fragments, thereby generating full-length genes (**Fig. 8**). In this case, the resulting fragments prime each other based on sequence homology, and recombination occurs when fragments from a gene prime those from another gene copy.

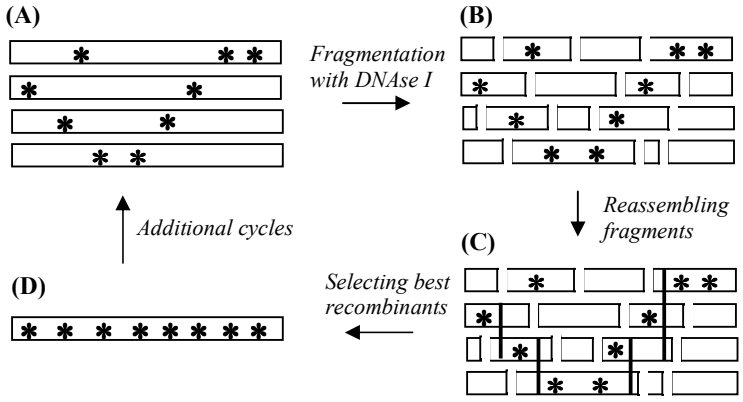

Fig. 8. Enzyme engineering by DNA shuffling. A pool of homologous genes harbouring various point mutations is digested by using DNAse I **(A)**, resulting in many fragments **(B)**. A PCR reaction is employed to let recombination occurs among the homologous fragments **(C)**, thereby generating a library of recombinant genes **(D)**. The recombinant gene with a desired property is screened from the library, and can be subject to further rounds of recombination to improve the property. Copied and modified from Stemmer, 1994.

By using DNA shuffling, genes from multiple parents and even from different species can be bred each other in a single step to create new hybrid genes (Arnold, 2001). For such reason, this method is considered as an extremely powerful way for the molecular engineering of a protein or biocatalyst, generating protein mutants which are not found in nature. The advantage of this method over other existing mutagenesis methods are likely to increase the numbers of cycles of molecular evolution (Stemmer, 1994).

CONCLUDING REMARKS

Biocatalysts originated from marine microorganisms have attracted much scientific attention for being used in various industrial processes. However, for many years the search for new biocatalysts of microbial origin was based on cultivation-dependent methods which count only less than 1% of all microorganisms in most environments, including marine environment.

A recent breakthrough in Microbial Ecology has led to the birth of Metagenomics, a powerful cultivation independent approach for gaining access to the enormous uncultured microbial diversity and their unlimited pool of new biocatalysts. With recent advances in DNA sequencing technology and bioinformatics tools, this approach have expanded into the sequence-based analysis of biocatalysts-encoding genes and phylogenetic markers. The results of such analysis can give useful information on the functional and structural properties as well as microbial source of the target biocatalysts. In spite of the potential of finding new biocatalysts from uncultured marine microorganisms, it often appears that natural biocatalysts need to be improved at molecular level to meet industrial and biotechnological requirements through protein engineering.

REFERENCES

Arnold, F.H., 2001. Combinatorial and Computational Challenges for Biocatalyst Design. *Nature* 409:253-258.

Baneyx, F., 1999. Recombinant Protein Expression in *Escherichia coli*. *Curr. Opin. Biotechnol.*10:411-421.

Bertoldo, C., G. Antranikian, 2002. Starch hydrolyzing enzymes from thermophilic archea and bacteria. *Curr. Opin. Chem. Biol*. *6*:151-160.

Bornscheuer, U.T. and M. Pohl, 2001. Improved Biocatalysts by Directed Evolution and Rational Protein Design. *Curr. Opin. Chem. Biol.* 5:137-143.

Campbell B.J., J.L. Stein, S.C. Cary. 2003. Evidence of chemolithoautotrophy in the bacterial community associated with *Alvinella pompejana*, a hydrothermal vent polychaete. *Appl. Environ. Microbiol.* 69: 5070–5078

Cottrel, M.T., J.A. Moore, and D.L. Kirchman, 1999. Chitinases from Uncultured Marine Micro-organisms. *Appl. Environ. Microbiol.* 65(6):2553-2557.

Cottrell, M.T., D.N. Wood, L. Yu, and D.L. Kirchman, 2000. Selected Chitinases Genes in Cultured and Uncultured Marine Bacteria in the α- and γ-subclasses of the Proteobacteria. *Appl. Environ. Microbiol.* 66(3):1195-1201.

Contursi, P., R. Cannio, S. Prato, G. Fiorentino, M. Rossi, and S. Bartolucci. 2003. Development of a genetic system for hyperthermophilic archaea: expression of a moderate thermophilic bacterial alcohol dehydrogenase gene in *Sulfolobus solfataricus*. *FEMS Microbiol. Lett.* 218:115-120.

Courtois, S., Cappellano C.M., Ball M., Francou F-X, Normand, Helynck G., Martinez A., Kolvek S.J., Hopke J., Osburne M.S., August P.R., Nalin R., Guerineau M., Jeannin P., Simonet P., and Pernodet J-L., 2003. Recombinant Environmental Libraries Provide Access to Microbial Diversity for Drug Discovery from Natural Products. *Appl. Environ. Microbiol.* 69(1):49-55.

Daniel, R. 2001. Environmental genomics: construction, sequencing, and screening of environmental libraries with respect to enzymes and drugs of industrial relevance, in *Competence network genome research on bacteria for the analysis of biodiversity and its further use for the development of new product processes*. Editors: G. Gottschalk, D. Trezeciok, and K. Hahne. The institute of Microbiology, University of Göttingen, Germany

Estell, D.A., T.P. Graycar, and J.A. Wells, 1985. Engineering an Enzyme by Site-Directed Mutagenesis to be Resistant to Chemical Oxidation. *J. Biol. Chem.* 260(11):6518-6521.

Ginolhac, A., C. Jarrin, B. Gillet, P. Robe, P. Pujic, K. Tuphile, H. Bertrand, T.M. Vogel, G. Perriere, P. Simonet, and R. Nalin. 2004. Phylogenetic Analysis of Polyketide Synthase I Domain from Soil Metagenomic Libraries Allows Selecetion of Promisisng Clones. *Appl. Environ. Microbiol.* 70(9):5522-5527.

Hallam S.J., P.R. Girguis, C.M. Preston, P.M. Richardson, E.F. DeLong. 2003. Identification of methyl coenzyme M reductase A (*mcr* A) genes associated with methane-oxidizing archaea. *Appl. Environ. Microbiol.* 69:5483–91.

Handelsman J., 2004. Metagenomics: Application of Genomics to Uncultured microorganisms. *Microbiol. Mol. Biol. Rev.* 68(4):669-685.

Henne, A., R.A. Schimitz, M. Bomeke, G. Gottschalk, and R. Daniel. 2000. Screening of environmental DNA libraries for the presence of genes

conferring lipolytic activity on *Escherichia coli*. *Appl. Environ. Microbiol.* 66:3113-3116.

Hughes D.S., H. Felbeck, J.L. Stein. 1997. A histidine protein kinase homolog from the endosymbiont of the hydrothermal vent tubeworm *Riftia pachyptila*. *Appl. Environ. Microbiol.* 63:3494–3498

Jha, R. K. and X. Zi-rong, 2004. Biomedical Compounds from Marine Organisms. *Mar. Drugs 2:123-146.*

Knietsch A., T. Waschkowitz, S. Bowien, A. Henne, and R. Daniel, 2003. Construction and Screening of Metagenomic Libraries Derived from Enrichment Cultures: Generation of a Gene Bank for Genes Confferring Alcohol Oxidoreductase Activity on *Escherichia coli*. *Appl. Environ. Microbiol.* 69(3):1408-1416.

Lucas, S., L. Toffin, Y. Zivanovic, D. Charlier, H. Moussard, P. Forterre, D. Priuer, and G. Erauso. 2002. Construction of a shuttle vector for, and spheroplast transformation of, the hyperthermophilic archaeon *Pyrococcus abysii*. *Appl. Environ. Microbiol.* 68(11):2258-5536.

Machielsen R., A.R. Uria, S.W.M. Kengen, and J. van der Oost, 2006. Production and Characterization of a Thermostable Alcohol Dehydrogenase That belongs to the Aldo-keto Reductase Superfamily. *Appl. Environ. Microbiol.* 72(1):233-238.

Madigan, M.T., J.M. Martinko, and J. Parker. 2000. Brock Biology of Mircoorganisms. Ninth Edition. *Prentice-Hall, Inc*. Upper Saddle River, New Jersey, USA. p 165-167.

Novagen 2002-2003 Catalog. Protein Expression: pET System Tutorial. www.novagen.com.

Piel J., Hui D., Wen G., Butzke D., Platzer M., Fusetani N., Matsunaga S., 2004. Antitumor Polyketide Biosyntesis by an Uncultivated Bacterial Symbiont of the Marine Sponge, *Theonella swinhoei. PNAS* 101(46):16222-16227.

Prescott, L.M., J.P. Harley, and D.A. Klein, 2002. *Microbiology*. Fifth edition. The McGraw-Hill Company, Inc. New York, USA. 1022 p.

Reymond, J-L., 2001. New High-Throughput Screening Assays for Biocatalysis. *Chimia* 55(12):1049-1052.

Riesenfeld, C.S., P.D. Schloss, and J. Handelsman, 2004. Metagenomics: Genomic Analysis of Microbial communities.*Annu. Rev. Genet.* 38:525-553.

Rondon, M.R., P.R. August, A.D. Betterman, 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, and R.M. Goodman, 2000. Cloning the Soil Metagenome: a Strategy for Accessing the Genetic and Functional Diversity of Uncultured Microorganisms. *Appl. Environ. Microbiol.* 66(6):2541-2547.

Schleper C, E.F. DeLong, C.M. Preston, R.A Feldman, K.Y. Wu. 1998. Genomic analysis reveals chromosomal variation in natural populations of the uncultured psychrophilic archaeon *Cenarchaeum symbiosum*. *J. Bacteriol.* 180: 5003–5009

Schut, G.J., S.D. Brehm, S. Datta, M.W.W. Adams. 2003. Whole-genome microarray analysis of a hyperthermophile and an archaeon: *Pyrococcus furiosus* grown on carbohydrates or peptides. *J. Bacteriol.* 185:3935-3947.

Schloss, P.D. and J. Handelsman, 2003. Biotechnological Prospects from

Metagenomics. *Curr. Opin. Biotechnol.* 14:303-310.

Stemmer, W.P.C., 1994. DNA Shuffling by Random Fragmentation and Reassembly: *In Vitro* Recombination for Molecular Evolution. *Proc. Natl. Acad. Sci.* 91:10747-10751.

Suzuki, M.T., M.S. Rappe, Z.W. Haimberger, H. Winfiled, N. Adair, J. Strobel and S. Giovannoni, 1997. Bacterial Diversity among Small-Subunit rRNA Gene Clones and Cellular Isolates from the Same Seawater Sample. *Appl. Environ. Microbiol.* 63(3): 983-989.

Tanner, M.A., C.L. Everett, W.J. Coleman, M.M. Yang, and D.C. Youvan, 2000. Complex Microbial Communities Inhabiting Sulfide-rich Black Mud from Marine Coastal Environments. *Biotechnol. et alia* 8:1-16.

Uria, A.R., Y.N. Fawzya, dan E. Chasanah, 2004. Genomika Fungsional Kelautan: Prospeknya dalam mengali potensi molekuler organisme laut (Marine Functional Genomics: its Prospective in the molecular exploitation of marine organisms). *Warta Penelitian Perikanan Indonesia. Edisi Pasca Panen dan Sosial Ekonomi* 10(7):17-22. (in Indonesian)

Uria, A.R., 2004. Purification and Characterization of Recombinant NAD(P)-dependent Alcohol Dehydrogenase from the Marine Hyperthermophilic Archaebacte- rium, *Pyrococcus furious* and Generation of a DNA construct for its Overexpression in the Deep-sea Hyperthermophile, *P. abysii*. *A Seven-month Research Thesis of the M.Sc Programme in Biotechnology*, Wageningen University and Research Centre, the Netherlands.

Van der Oost, J. and L. de Graaf, 2002. Applied Molecular Genetics. *A Teaching Module*. Wageningen University and Research Centre, the Netherlands.

Van den Burg, B., G. Vriend, O.R. Veltman, G. Venema, and V.G.H. Eijsink, 1998. Engineering an Enzyme to Resist Boiling. *Proc. Natl. Acad. Sci.* 95:2056-2060.

Wahler, D. and J-L Reymond, 2001. Novel Methods for Biocatalyst Screening. *Curr. Opin. Biotechnol.* 5:152-158.

Wahler, D. and J-L Reymond, 2001. High-throughput Screening for Biocatalysts. *Curr. Opin. Biotechnol.* 12:535-544.