Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Molecular Catalysis B: Enzymatic 26 (2003) 241–250
JOURNAL OF MOLECULAR CATALYSIS B: ENZYMATIC
www.elsevier.com/locate/molcatb

Expanding functional spaces of enzymes by utilizing whole genome treasure for library construction

Hyo-Jeong Oh, Kyoung-Won Cho, In-Su Jung, Won-Ho Kim, Byung-Ki Hur, Geun-Joong Kim*

College of Engineering, Institute of Biotechnological Industry, Inha University, 253, Yonghyun-dong, Nam-gu, Incheon 402-751, South Korea

Received 18 March 2003; received in revised form 26 June 2003; accepted 30 June 2003

We pay a tribute to Professor Joon-Shick Rhee on the occasion of his retirement from the position of professor in the Department of Biological Sciences, Korea Advanced Institute of Science and Technology. His extraordinary academic career has been spanning nearly three decades. He is universally admired and respected for his crucial contributions to the research field of enzyme engineering, applied microbiology and food science

Abstract

A huge database resulted from whole genome sequencings has provided a possibility of new information that is likely to extent the scope and thus changes the way of approach for the functional assigning of putative open reading frames annotated by whole genome sequence analyses. These are mainly realized by ease, one-step identification of putative genes using genomics or proteomics tools. A major challenge remained in biotechnology may translate these informations into better ways to screen or select a gene as a representative sequence. Further attempts to mine the related whole genes or partial DNA fragments from whole genome treasure, and then the incorporation of these sequences into a representative template, will result in the use of genetic information that can be translated into functional proteins or allowed the generation of new lineages as a valuable pool. Such screens enable rapid biochemical analysis and easy isolation of the target activity, thereby accelerating the screening of novel enzymes from the expanded library with related sequences. Information-based PCR amplification of whole genes and reconstitution of functional DNA fragments will provide a platform for expanding the functional spaces of potential enzymes, especially when used mixed- and metagenome as gene resources.
© 2003 Elsevier B.V. All rights reserved.

Keywords: Enzyme library; Functional space; Whole genome; Metagenome; FSS

1. Introduction

Biocatalyst-mediated, mainly by enzymes, various reactions are attractive for a lot of purposes due to their efficiency and selectivity in the reaction chemistry [1]. Functionally precise enzymes have a good performance and their versatilities provide an extremely diverse pool of activities or functional clues for tailoring the enzyme by protein engineering, thereby well-matching with most reactions found naturally or exploited industrially [2]. In this context, naturally occurring enzymes steadily attract attentions

* Corresponding author. Tel.: +82-32-860-7512; fax: +82-32-872-4046.
E-mail address: geunkim@inha.ac.kr (G.-J. Kim).

1381-1177/$ – see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/j.molcatb.2003.06.007

for finding a known or new activity. Genetic materials encoding either intact open reading frames or their fragments, therefore, are being pursued to satisfy the increasing demands for new biocatalysts, which focus on novel functions to break current barriers [3]. Although various approaches have been continued to annotate protein function in vitro, or in vivo, and thus used to screen the potential enzymes with new functions, a challenge still be remained because the searching spaces are mainly within whole cell enzymes enriched from natural niches suspected [4].

The established screening or selection methods, in general, find a candidate from a pool of enzymes that constitutively expressed or high-levelly induced in enriched conditions. Therefore, lower activities in screening step do not mean that such enzyme has a relatively low potential, because various factors can not permit all enzymes to be expressed equally due to tight regulation or repression in vivo. Therefore, for the selection of indeed potential enzymes, it is ideally necessary to express all related enzymes as a library format, or, at least, genetic information encoding the relevant sequences is strictly required prior to screening the activity from natural resources [5,6]. Fortunately, with the advent of high-throughput molecular biology, it is now possible, within weeks, to assign responsible genes as representative sequences for library construction. This is currently an easy step, because the genome projects of more than 70 strains have been completed to date and a lot of draft sequences are also available in the projects progressed [7,8]. Thus, if a strain is chosen as a possible source for an activity, a responsible or plausible sequence is readily identified or deduced from its own or related genomes due to the conservative evolution of whole or consensus regions through the evolutionary procedure. These genome-based screens and bioinformatic analyses open new windows for the screening of novel biocatalysts and then library construction [5,9].

We here summarize a systematic approach that expand functional spaces of enzymes by combination of preexisting tools in screening and engineering steps of potential enzymes. The principles and applicable strategies of this approach are discussed briefly, based on previous reports that successfully applied the approach for practical cases [10,11].

2. Selection of representative sequences by information-based whole genome approach

2.1. Basic principle

An organized systematic approach for the selection of representative sequences from whole genome sequences is consisted of the following steps (Fig. 1). (1) Strain pools are enriched from natural niches by

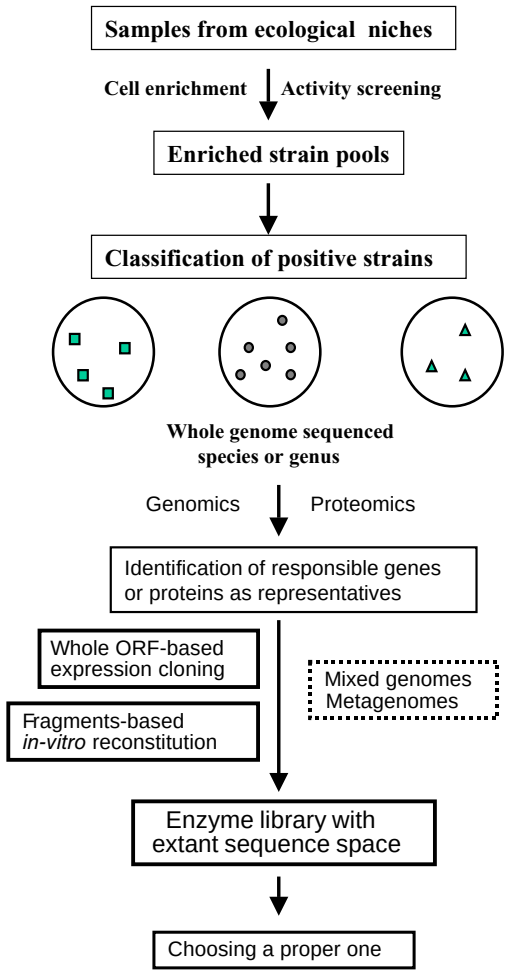

Fig. 1. The summarized typical procedures for the selection of potential enzyme pool. In this procedure, cell enrichment and activity screening provide a pool of diverse strains from nature. After strain identification, a strain pool is chosen and then analyzed for the properties of traced enzyme at protein level. Gathered information will be a basis for the mining of probable gene(s) from gene or genome databases.

typical procedures and analyzed for their potential with whole cells or crude extracts, which can result in a pool of strains as candidates. (2) The candidate strains are taxonomically identified, especially using strain- or genus-specific molecular marker(s). Among these strains identified, a whole genome sequenced strain is chosen as a source for the selection of a representative gene. (3) The traced activity as the target is proven at protein level by peptide mass fingerprinting, N-terminal- or internal protein sequencing after 1 or 2D gel electrophoresis of fractionated enzyme solution, thus a representative gene is readily selected from predetermined ORFs by whole genome sequencing. (4) After functional identification of the selected gene, related sequences are searched on protein and gene data bank, thereby yielding a relevant gene family. Based on this information, various sources of genomes from cultivable and uncultivable strains (metagenome) are used for expanding functional spaces of the representative by information-based PCR cloning and functional reconstitution. In step (1), all cultivable cells, if possible, should be enriched for further analyses. Step (2) uses current techniques, such as RAPD, RFLP and RDA, which can classify reasonably the positive strains suspected. Along with this, direct protein sequencing tools permit an easy identification of a responsible gene. As for general use, the critical point in summarized procedure lies in designing of a simple and rapid detection method of the activity in either solution or solid culture.

2.2. Strain enrichment and activity screening

For activity screening, the cells, in general, are enriched in typical or selective media that optimally formulated for a high cell growth and protein expression, by considering various culture conditions [12,13]. From diverse ecological niches, an aliquot is sampled and then suspended in a specified medium or saline buffer solution. In this step, solution or solid cultures are incubated at various temperatures for appropriate times to maximize the enriched strain pools. According to the expected strain pools, the specific chemicals that inhibit selectively to a genus or had a broad range of inhibitory spectrum are also included [14]. Using well-isolated colonies under enriched conditions, strain pools are analyzed for their activity using a solution or solid culture, according to the sensitivity and procedure of enzyme assay. Primary selection or screening is also performed in an enriched medium supplemented with a substrate as an inducer. The addition time and interval to the medium strictly depend on the inhibitory effect of an inducer to cell growth.

In order to select potential enzymes with high selectivity or activity to the target chemical, all suspected strains should be compared repeatedly, and a tedious, but confirmatory, biochemical assay using whole cell enzymes or crude extracts must be incorporated into the final consideration.

2.3. Identification and classification of enriched strains

The next step is further analyses in the phenotypic and genotypic characters for strain classification into a genus (Fig. 2). With a pool of selected strains, the apparent phenotypes, such as a motility, Gram staining, fluorescence emission and oxygen demand, are considerable as basic properties. More specific traits, including an enzyme activity and carbon source utilization, are also included to be criteria. The optimal growth temperature and pH, as well as fatty acid composition of cell wall, may be a considerable factor.

At the molecular level, some techniques, including ribosomal DNA analyses, can provide a rapid protocol as the determinant. First, the DNA fingerprinting technique, random amplified polymorphic DNA (RAPD), is used as the most sensitive method for distinguishing different strains within natural isolates. RAPD utilizes a single short primer of arbitrary sequence in a reaction of PCR [15]. The amplified DNA fragments are subjected to gel electrophoresis, and then analyzed for the patterns and sizes as determinants. For reproducible results, it is vital that the template DNA can be prepared by a consistent and reliable method for high purity, reducing a level of contaminants that inhibit PCR or degrade the template DNA. In addition, the quantity of template DNA is also an important criterion for establishing reliable RAPD patterns [16]. Although PCR for RAPD is performed generally under standard conditions, in some cases, annealing at a low stringency (<45 °C) for an appropriate cycle and the subsequent annealing at a high stringency (45–60 °C) for remaining cycles are proposed to reduce nonspecific amplification. The resulting DNA fragments are,

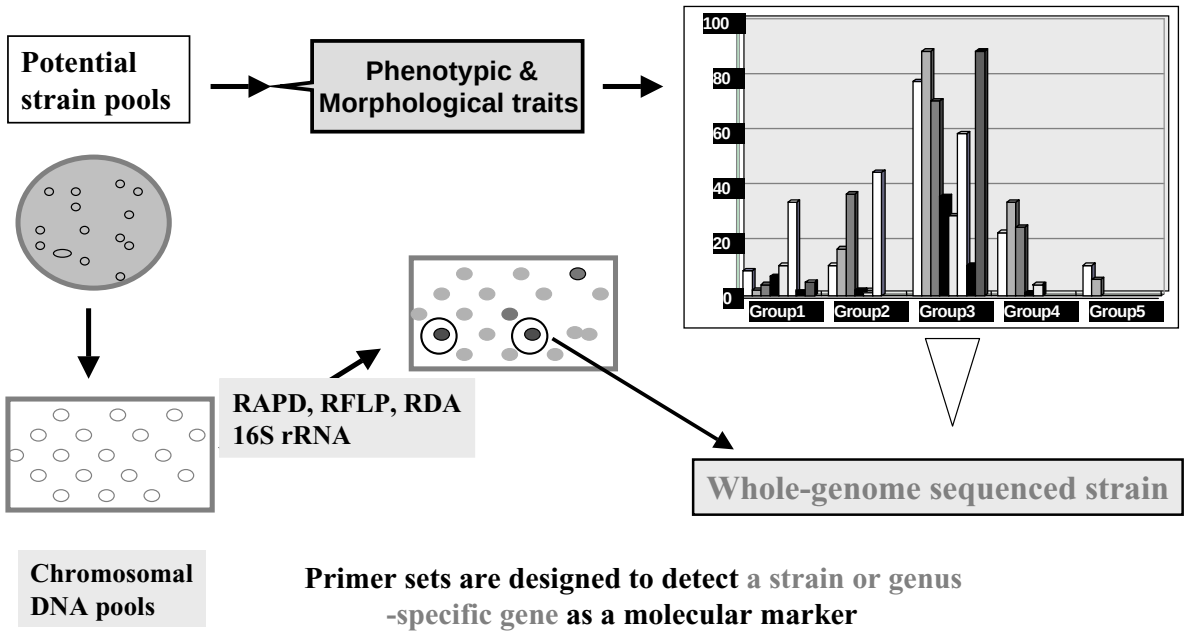

Fig. 2. The classification step of enriched strains. Both biochemical and molecular biological markers are all considerable as indicators. Critical determinants are readily prepared for cell sorting by some techniques, such as RAPD, RFLP and RDA, using prevalent tools PCR and hybridization.

occasionally, used for the template of other technique, termed RFLP (restriction fragment length polymorphism). This combined protocol can ensure the result to identify the strains precisely [17].

The technique, termed multiplex PCR with mixed primers spanning strain- or genus-specific genes, can also be utilized for strain identification. This comprises the successive step composed of PCR amplification and proving the specific band by hybridization or resulting sizes [18], and currently provides a method that differentiates the presence of a species in isolates. Another approach that utilizes one of a variety of subtractive techniques to recover genes present in one isolate but not the other is also proposed to identify a strain in isolates. Such technique, termed representational difference analysis (RDA), can be adapted as a way for identifying a genome of related strains or different genus [19].

Among the classified strains, either whole-genome sequenced strain(s) or closely related strain to a whole genome sequenced strain, is chosen for the selection of a representative sequence as the template for library construction.

2.4. Selection of a representative gene

The subsequent step is direct and simple identification of a responsible gene in the chosen strain by analyzing the property at protein level (Fig. 3). For the purpose, the identical, or closely related ones, to the chosen strain is obtained from culture collection as positive controls. The chosen candidate is first analyzed as a possible source for a representative gene, by activity staining (if possible) on native PAGE using crude extracts. If it is impossible to assign the activity to a resolved band on PAGE, there is a need for separating the protein band from other contaminants. This fractionation is also required to exclude the possibility that more than an enzyme could act on the identical substrate, due to the shared substrate spectrum When a distinct band corresponding to the expected enzyme is appeared in a fraction, the resulting solution can be used for protein identification. The critical information for gene mining must be resulted from the following steps of protein sequencing.

The N-terminal sequence analysis of a protein is a prevalent tool performed according to the general

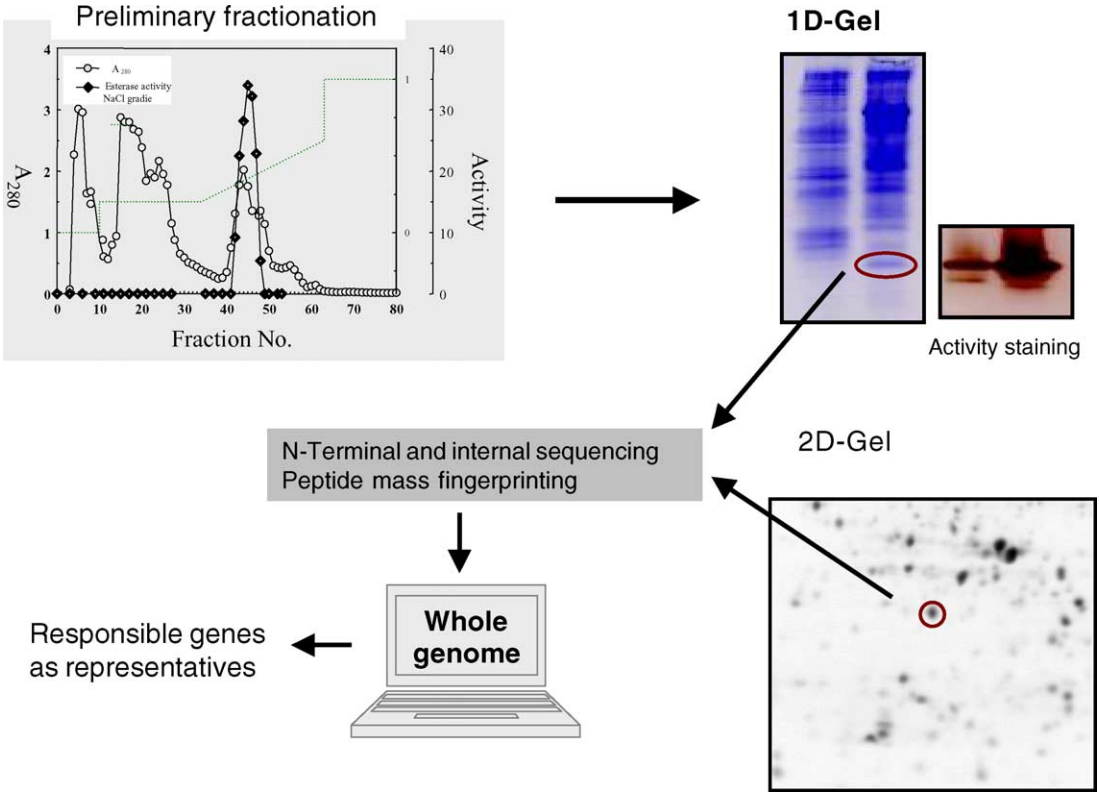

Fig. 3. Rapid identification of a protein as the template for library construction. The traced enzyme as the target is assayed and then identified by protein sequencing tools, such as peptide mass fingerprinting, N-terminal and internal protein sequencing, using MS spectrophotometer after a gel electrophoresis. Therefore, a representative gene is readily selected from open reading frames predicted from whole genome sequence analyses.

protocol of the Edman degradation [20]. The protein sample runs on PAGE and transblots onto polyvinylidene difluoride (PVDF) membrane, and then sliced with a resolved band for sequencing. Currently, 20–25 residues can be assigned accurately by using about 10–30 pmol of proteins or peptides. This technique allows us to directly analyze the peptide sequence of the target protein from a complex mixture. A modified technique that uses the passively eluted sample from a sliced gel is directly applied to the sequencer cartridge without the need for the protein blotting onto PVDF membrane [21]. Direct sequencing of peptides bound to the synthetic or immobilized resin is also possible [22].

Internal protein sequencing is one of the most useful techniques to get amino acid sequences from N-terminally blocked or complex proteins [23]. In this procedure, the protein is first subjected to proteolysis that cleaves the protein into peptides of different lengths. After separation step using HPLC or other compatible tools, the resulting peptide(s) is applied to the typical step of N-terminal sequencing. Alternatively, peptide mass fingerprinting is also possible in many cases to identify a protein by subjecting the digested peptides, either separated or not from reaction mixture, to nanospray mass spectrometry [24]. It is also possible to obtain additional sequence information by fragmenting the peptide mass in a technique called ESI-MS/MS used in conjunction with tandem mass spectrometry [25]. The obtained mass is compared to the calculated molecular mass of annotated sequences for gene identification.

With current tools, the scan on a whole genome sequence for gene mining can be succeed even with

partial sequences of protein. In this step, the perfect assignment (∼100%) of genes to their activity by technical combination between PAGE and protein sequencing is possible if ORF is predicted accurately from whole genome sequencing.

2.5. Functional identification and database search for structural information

To verify the encoded activity, the next step is a PCR-amplification of the mined gene from genomic DNA of a whole genome sequenced strain. When performed with a related strain and degenerated primers, a low stringent PCR is also applied to enhance the amplification of related genes. The amplified genes are cloned and analyzed for protein expression and activity. It is notify that the cloned gene as a representative does not mean that its encoded enzyme has a plentiful activity, thus purification of the expressed enzyme to apparent homogeneity is occasionally required. To avoid the risk that causes different reading frames by a low stringent PCR with degenerated primers, the amplified genes are also cloned and expressed in other vectors, utilizing a different reading frame. Analyses in terms of solubility, localization and expression level will be a basis for further selection or engineering. As a criterion for successful mining, crude extracts or purified enzymes are analyzed for the tracing activity using the target chemical and its derivatives.

To search the related family for library construction, the representative sequence is compared automatically on annotated sequences of protein pools from various gene and protein data banks. Currently, the GenBank database is the most well known one accessible freely through NCBI [26]. Data submission and query are linked with a retrieval system Entrez, which integrates data from the major DNA and protein sequence databases, along with information on taxonomy, genome, protein structure and organization. The BLAST family of programs as searching tools for sequence similarity is steadily improved and provided the conserved positions in close or distant family members [27], which enables us to select better positions that may play key roles in determining the related sequences.

As an important hub for public access, the PIR web site also provides data mining and sequence analysis tools for finding related sequences, with functional information for submitted sequences [28]. This database is consisted of the three major databases, PSD, NREF and iProClass, all of which form a basis for providing the searching results with retrieval lists. These results are strictly dependent on sequence unique identifiers of all underlying databases (PIR, SWISS-PROT and RefSeq). When submitted a query sequence, it returns protein entries listed in summary lines with information on protein name including ID, matched field, taxonomy, superfamily, domain and motif [28]. A computer based sequence search and analysis protocol, DomainFinder, based on PSI-BLAST and IMPALA, has been developed for integrating gene sequences from GenBank into their respective structural families within the CATH domain database [29], thereby assigning a new sequence to a CATH homologous superfamily. As for further detailed analyses in substructural annotation, the structural classification of proteins (SCOP) and SUPFAM databases provide a comprehensive description of the relationships between known protein structures [30].

The resulting sequences with functional or structural relatedness are aligned, mainly, by hierarchical clustering of the individual sequences based on the pairwise similarity scores. The conserved pattern of amino acid residue in related genes is usually analyzed by Clustal W program [31] and then proofed by other similar programs. Subsequently, the conserved regions and patterns resulted from sequence alignment tools are analyzed again by BLAST and FASTA search, or related programs including phylogenetic tree analysis [32], hidden Markov model (HMMER) domain [33] and conserved domains (CDD) search [34].

The structure–function information and searching results for the related sequences will play key roles in further cloning and library construction to the subsequent step for expanding functional spaces. Available current tools and resources for library construction are partly unsatisfactory due to their narrow scope and limited sequence spaces, although considerable results are being reported [4]. Therefore, to broaden functional spaces, the mined sequences are further extended into new sequence spaces by utilizing limitless resources (mixed or metagenome) and novel approaches for protein engineering.

3. Utilization of metagenome as a source for new functional sequence spaces

As an emerging field for new biological (or functional) space, recent studies have revealed that only a tiny fraction of microbes in nature are accessed by traditional cultivation methods, thus almost all fraction (>99%) of microbes might be remained to be explored [35]. This problem is mainly due to the fact that enrichment or pure culture is traditionally preceded prior to the selection or screening of responsible genes for the expected activity. Attempts to investigate the full extent of microbial diversity are, therefore, conducted by using a bacterial artificial chromosome (BAC) or in vitro packaging cosmid system as cloning vectors for DNA pools (metagenome) of uncultivated strains [36,37]. However, an inherent shortcoming is existed in current vector systems that are unable to guarantee the expression and independency of cloned inserts. Therefore, information-based PCR cloning of whole ORF or minimum functional domain from metagenome is currently adapted as new routes, and the related fields will grow rapidly to access uncultured bacteria as genetic resources (Fig. 4). Using PCR with degenerated primers that spanned either a whole ORF or minimum functional domain, related DNA fragments are amplified and cloned into an expression vector, followed by analyzing the inserts as controls for further strategy. It is noted that the amplification of all possible relatives is a crucial factor for information-based library construction using metagenome. Therefore, a strategy using RNA to suppress the reamplification of known members of related family is presumed as a valuable tool [38].

The advent of powerful engineering tools, such as in vitro recombination using PCR for sexual recombination (DNA shuffling), domain reconstitution and back crossing, have ensured DNA fragments to serve as potential genetic materials, thereby broadening functional spaces into more diverse sequences [39]. Therefore, fragments (domain or region) based in vitro reconstitution, instead of whole ORF based approach, will appear dominantly in the related fields of metagenome (Fig. 5). During this procedure,

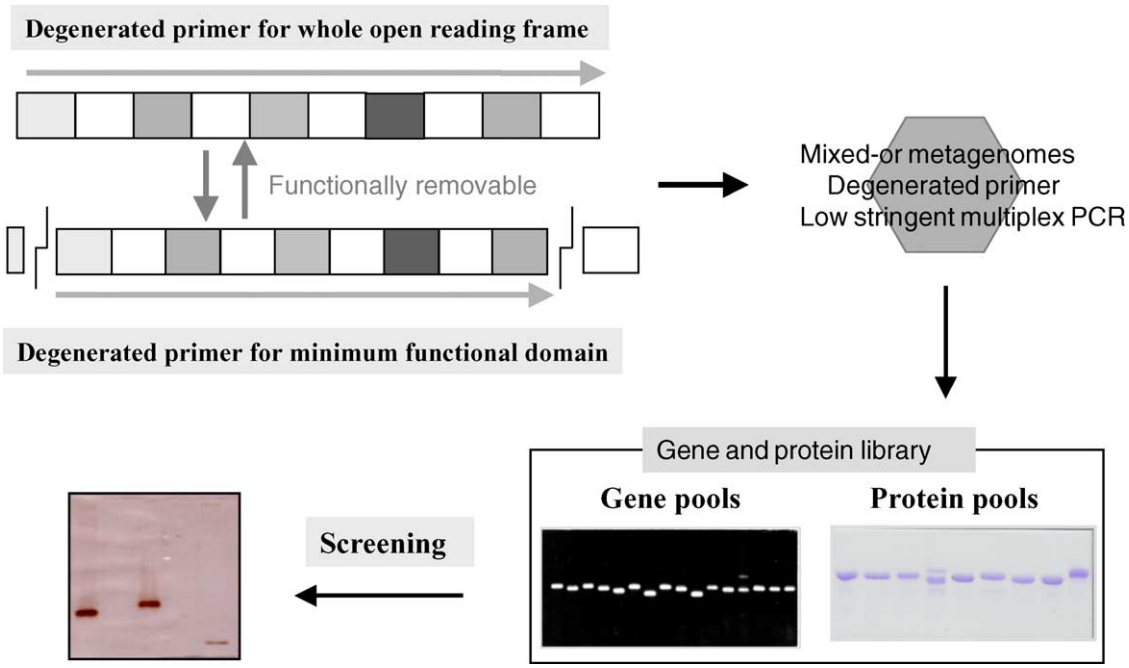

Fig. 4. Information-based expression cloning by PCR for library construction. After functional identification of gene(s) as representative(s), the related sequences are searched on protein and gene data banks. Based on this information, various sources of mixed or metagenome are employed as resources for new functional spaces. In this library, consistent information about whole ORF or minimum functional domains is guarantee the success of expression cloning.

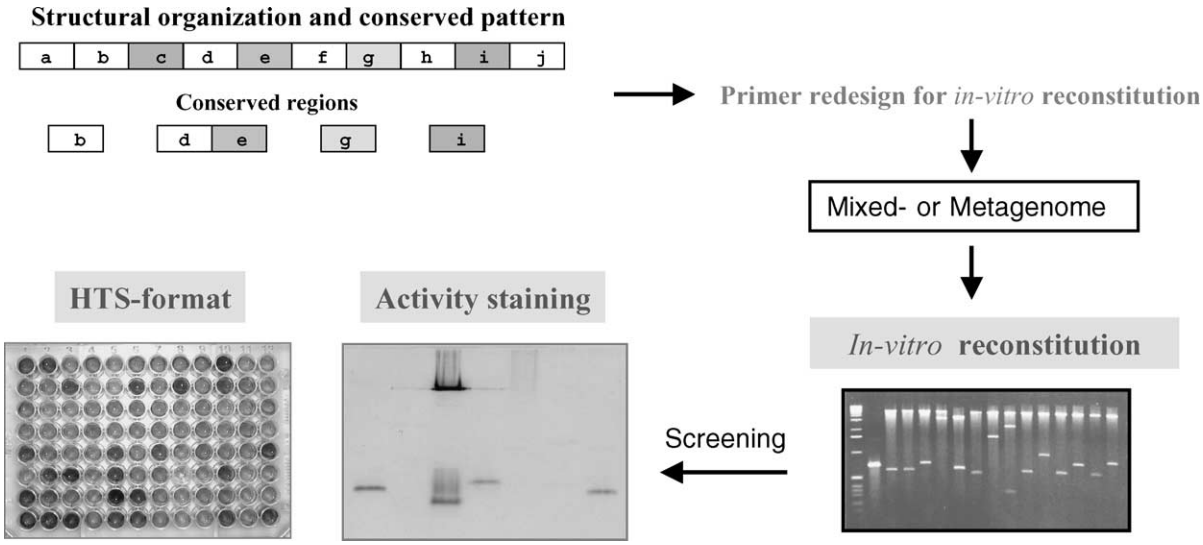

Fig. 5. Information-based in vitro reconstitution of functional enzymes for library construction. Based on the structural information about the conserved patterns, the related DNA fragments are amplified by PCR using degenerated primers from mixed or metagenome, and then reconstituted by sexual recombination (DNA shuffling) or domain swapping. During this procedure, reconstituted regions, domains and whole ORF occasionally can backcross with the original representative. The critical point in library construction lies in designing of a simple and rapid detection method of enzyme activity in either solution or solid culture, which guarantees the strategy for general use.

reconstituted genes occasionally can backcross with the original representative, concomitantly producing both novel gene fragments and engineered enzymes. This procedure should also adapt a strategy that specifically inhibited the reamplification of dominant gene fragments. The two independent approaches of both information- and activity-based screening tools will provide considerable ways to discover novel enzymes with functionally new sequence spaces, and will be necessary equally to access molecular treasures uncovered in nature. Preliminary steps for isolation and fractionation of complex chromosome mixtures primarily guarantee the successful applications of metagenome as potential resources.

4. Expanding functional spaces of representative sequences by FSS

Recent advances in protein engineering over prevalent tools have accelerated the understanding of a number of intrinsic questions regarding protein evolution in nature, and also provided an effective tool to generate the proteins with new sequence spaces of different functions. The new concepts are mainly based on the incorporation or deletion of random sequences into a terminus or internal region of the target gene, providing enzyme lineages that frequently occurred in vivo but scarcely in vitro [40,41]. In this context, a novel approach, termed functional salvage screen (FSS), to generate protein lineages with new sequence spaces through functional or structural salvage of a defective enzyme has been designed and evaluated for its potential [42]. As shown in Fig. 6, the FSS starts with a construction of the defective template expressing no activity by genetically disrupting a predetermined region(s). The defective template is designed to be unable to recover the function in vivo by simple insertion or deletion of base(s). Thus, only a recombination between a defective template in an arbitrary region(s) and gene segments derived from a pool of genomes, including metagenome, could rescue the protein function. These events are realized by the incorporation of a gene fragment through a sequence specific insertion resembling a shotgun cloning strategy [43] or homologous recombination using a PCR-like process [44]. For both cases, various pools of diverse genomic DNA that pretreated with either an

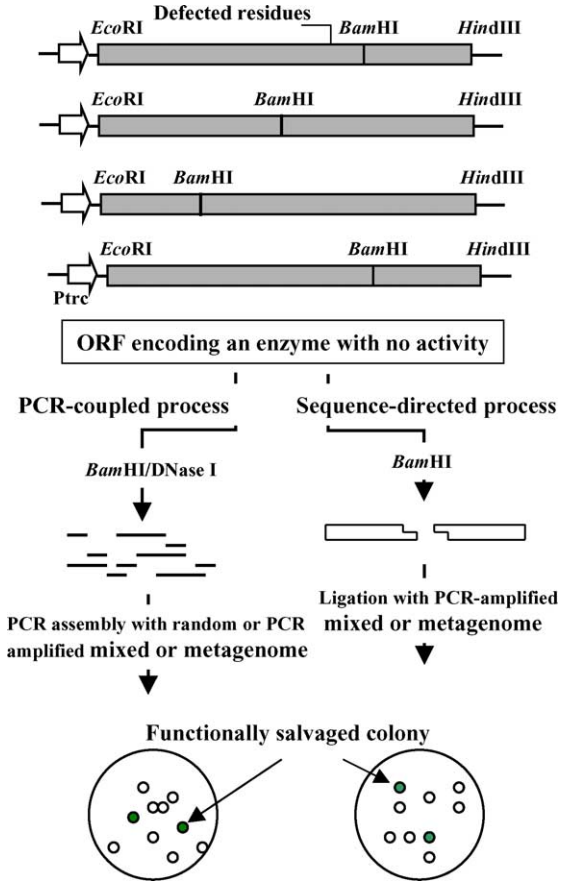

Fig. 6. Schematic diagram of the generation of new protein lineages by functional salvage process (FSS). The defective genes expressing no activity are first constructed by genetically disrupting a predetermined region(s) of the protein as starting templates. For the functional salvage process, two independent approaches, PCR-coupled and sequence-directed methods, are attempted to prepare new sequence- and/or functional spaces of enzyme for library construction.

enzyme (*Sau*3AI or DNaseI) or physical shearing are provided and reassembled with each (or mixed) of the defected template for FSS. They can also utilize randomly or specifically amplified pool of DNA for the identical purpose. As an interesting proposal, the defected variants will be expected to recover their function by complementing the defective gene at multiple sites. Therefore, more diverse protein lineages with lower homology, compared with the parent protein, are expected as the number of salvage points increase [42].

The FSS relies on the screening of the protein variants with appropriately reorganized structures complementing a defective trait. Either non-essential or essential regions of a protein should be the target sites for FSS, and these defects can be functionally and/or structurally complemented by utilizing huge genetic materials originated from any genetic resources, including metagenome. In this context, the FSS can be a valuable tool for the generation of an enzyme library with extreme diversity in sequence spaces, which may adapt a different evolutionary path from the parent enzyme. The resulting library can be further subjected to directed evolution for functional tuning.

5. Conclusions

A systematic approach for mining a representative sequence and then library construction is summarized here. Basically, the typical, but still important, processes for screening of enzymes from natural sources are preceded prior to further analyzing the suspected strain pool at protein level for the rapid identification of responsible genes. Either a low or high activity in a potential pool of strains should be the target for the representative sequence. After choosing a sequence as the representative, a number of programs and searching tools for data mining are well matched with our intention to gather related genes from huge databases. Thus, the subsequent steps eventually rely on bioinformatic tools, which can lead us to set the criteria for mining the related sequences. These steps may play essential roles in the identification and cloning of relevant genes or their fragments to the subsequent step of the library construction, by employing mixed or metagenome in information- and activity-based approaches. The incorporation of these resources with new functional spaces into a library can provide a platform for utilizing whole genome treasures.

Acknowledgements

This work was supported by the Korea Research Foundation Grant (KRF-2002-005-D00005). We also acknowledged financial support by the 21C Frontier Microbial Genomics and Applications Center

Program, Ministry of Science & Technology (grant MG02-0301-001-1-0-2).

References

[1] K.E. Jaeger, B.W. Dijkstra, M.T. Reetz, Annu. Rev. Microbiol. 53 (1999) 315.
[2] J. Ogawa, S. Shimizu, Trends Biotechnol. 17 (1999) 13.
[3] H. Dalboge, L. Lange, Trends Biotechnol. 16 (1998) 265.
[4] M.R. Rondon, P.R. August, A.D. Bettermann, S.F. Brady, T.H. Grossman, M.R. Liles, K.A. Loiacono, B.A. Lynch, I.A. MacNeil, C. Minor, C.L. Tiong, M. Gilman, M.S. Osburne, J. Clardy, J. Handelsman, R.M. Goodman, Appl. Environ. Microbiol. 66 (2000) 2541.
[5] M.R. Martzen, S.M. McCraith, S.L. Spinelli, F.M. Torres, S. Fields, E.J. Grayhack, E.M. Phizicky, Science 286 (1999) 1153.
[6] G. DeSantis, Z. Zhu, W.A. Greenberg, K. Wong, J. Chaplin, S.R. Hanson, B. Farwell, L.W. Nicholson, C.L. Rand, D.P. Weiner, D.E. Robertson, M.J. Burk, J. Am. Chem. Soc. 124 (2002) 9024.
[7] R.A. Clayton, O. White, C.M. Fraser, Curr. Opin. Microbiol. 5 (1998) 562.
[8] E.V. Koonin, M.Y. Galperin, Curr. Opin. Genet. Dev. 6 (1997) 757.
[9] M. Carlson, Trends Genet. 16 (2000) 49.
[10] G.S. Choi, J.Y. Kim, J.H. Kim, Y.W. Ryu, G.J. Kim, Protein Expr. Purif. 29 (2003) 85.
[11] J.Y. Kim, G.S. Choi, I.S. Jung, Y.W. Ryu, G.J. Kim, Protein Eng. 16 (2003) 354.
[12] P. Lorenz, K. Liebeton, F. Niehaus, J. Eck, Curr. Opin. Biotechnol. 13 (2002) 572.
[13] J.Y. Kim, G.S. Choi, Y.J. Kim, Y.W. Ryu, G.J. Kim, J. Mol. Catal. B Enzymatic 18 (2002) 133.
[14] R.J. Arko, T. Odugbemi, J. Clin. Microbiol. 20 (1984) 1.
[15] J. Welsh, M. McClelland, Nucleic Acids Res. 18 (1990) 7213.
[16] M.R. Micheli, R. Bova, E. Pascale, E. D'Ambrosio, Nucleic Acids Res. 22 (1994) 1921.
[17] S.S. Salimath, A.C. de Oliveira, I.D. Godwin, J.L. Bennetzen, Genome 38 (1995) 757.
[18] S.N. Bourque, J.R. Valero, J. Mercier, M.C. Lavoie, R.C. Levesque, Appl. Environ. Microbiol. 59 (1993) 523.7.
[19] N. Lisitsyn, N. Lisitsyn, M. Wigler, Science 259 (1993) 946.
[20] E.L. Cannon, R.E. Lovins, Anal. Biochem. 46 (1972) 33.
[21] Z.H. Huang, T. Shen, J. Wu, D.A. Gage, J.T. Watson, Anal. Biochem. 268 (1999) 305.
[22] C.G. Fields, V.L. VanDrisse, G.B. Fields, Pept. Res. 6 (1993) 39.
[23] C.P. Vogt, A. Willi, D. Hess, P.E. Hunziker, Electrophoresis 17 (1996) 892.
[24] P. James, M. Quadroni, E. Carafoli, G. Gonnet, Protein Sci. 8 (1994) 1347.
[25] R.D. Smith, J.A. Loo, C.G. Edmonds, C.J. Barinaga, H.R. Udseth, Anal. Chem. 62 (1990) 882.
[26] D.A. Benson, I. Karsch-Mizrachi, D.J. Lipman, J. Ostell, B.A. Rapp, D.L. Wheeler, Nucleic Acids Res. 30 (2002) 17.
[27] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, J. Mol. Biol. 215 (1990) 403.
[28] C.H. Wu, H. Huang, L. Arminski, J. Castro-Alvear, Y. Chen, Z.Z. Hu, R.S. Ledley, K.C. Lewis, H.W. Mewes, B.C. Orcutt, B.E. Suzek, A. Tsugita, C.R. Vinayaka, L.S. Yeh, J. Zhang, W.C. Barker, Nucleic Acids Res. 30 (2002) 35.
[29] F.M. Pearl, C.F. Bennett, J.E. Bray, A.P. Harrison, N. Martin, A. Shepherd, I. Sillitoe, J. Thornton, C.A. Orengo, Nucleic Acids Res. 31 (2003) 452.
[30] S.B. Pandit, D. Gosar, S. Abhiman, S. Sujatha, S.S. Dixit, N.S. Mhatre, R. Sowdhamini, N. Srinivasan, Nucleic Acids Res. 30 (2002) 289.
[31] J.D. Thompson, D.G. Higgins, T.J. Gibson, Nucleic Acids Res. 22 (1994) 4673.
[32] S. Balaji, S. Sujatha, S.S. Kumar, N. Srinivasan, Nucleic Acids Res. 29 (2001) 61.
[33] A. Bateman, E. Birney, R. Durbin, S.R. Eddy, R.D. Finn, E.L. Sonnhammer, Nucleic Acids Res. 27 (1999) 260.
[34] A. Marchler-Bauer, J.B. Anderson, C. DeWeese-Scott, N.D. Fedorova, L.Y. Geer, S. He, D.I. Hurwitz, J.D. Jackson, A.R. Jacobs, C.J. Lanczycki, C.A. Liebert, C. Liu, T. Madej, G.H. Marchler, R. Mazumder, A.N. Nikolskaya, A.R. Panchenko, B.S. Rao, B.A. Shoemaker, V. Simonyan, J.S. Song, R.A. Thiessen, S. Vasudevan, Y. Wang, R.A. Yamashita, J.J. Yin, S.H. Bryant, Nucleic Acids Res. 31 (2003) 383.
[35] P. Lorenz, C. Schleper, J. Mol. Catal. B Enzymatic 19 (2002) 13.
[36] J. Handelsman, M.R. Rondon, S.F. Brady, J. Clardy, R.M. Goodman, Chem. Biol. 10 (1998) R245.
[37] K.T. Seow, G. Meurer, M. Gerlitz, E. Wendt-Pienkowski, C.R. Hutchinson, J.J. Davies, J. Bacteriol. 179 (1997) 7360.
[38] P.S. Yuen, K.M. Brooks, Y. Li, Nucleic Acids Res. 29 (2001) E31.
[39] W.P. Stemmer, Nature 370 (1994) 389.
[40] T. Matsuura, K. Miyai, S. Trakulnaleamsai, T. Yomo, Y. Shima, S. Miki, K. Yamamoto, I. Urabe, Nat. Biotechnol. 17 (1999) 58.
[41] A.E. Nixon, M. Ostermeier, S.J. Benkovic, Trends Biotechnol. 16 (1998) 258.
[42] G.J. Kim, Y.H. Cheon, M.S. Park, H.S. Park, H.S. Kim, Protein Eng. 14 (2001) 647.
[43] R. Jappelli, S. Brenner, Biochem. Biophys. Res. Commun. 266 (1999) 243.
[44] M. Kikuchi, K. Ohnishi, S. Harayama, Gene 236 (1999) 159.