FUNCTIONAL DIVERSITY OF BIOCATALYSTS: WHETHER TO BIOPROSPECT OR BIOENGINEER?

FARAH FADWA BENBELGACEM[1], OUALID ABDELKADER BELLAG[1], FATEMEH SOROODI[2], ABDUL AZIZ AHMAD[3], HAMZAH MOHD SALLEH[2], IBRAHIM ALI NOORBATCHA[1*]

[1]*Department of Biotechnology Engineering, Faculty of Engineering, International Islamic University Malaysia, Kuala Lumpur, Malaysia*
[2]*International Institute for Halal Research and Training (INHART), International Islamic University Malaysia, Kuala Lumpur, Malaysia*
[3]*Department of Science in Engineering, International Islamic University Malaysia, Kuala Lumpur, Malaysia*
**Corresponding author:ibrahiman@iium.edu.my*

ABSTRACT: Biocatalyst should have sufficient and efficient activity for the intended biotechnological application. In the quest for novel biocatalyst, there is a need to have a genetic diversity either by finding it within the astronomically large number of possible candidates or to obtain it by bioengineering an existing gene supported by various bioinformatic and molecular engineering tools. Nowadays, it is well-known that a huge number of microorganisms is unculturable and poses great challenges to access biocatalysts from these microbes. Metagenomics is one of the methods widely applied to reach out maximum possible variants to "bioprospect" biocatalysts. On the other hand, other approaches are available to bioengineer enzymes by modifying the DNA sequence precisely based on the structure and the function information of the protein in the case of rational design, or by a brave creation of anarchic mutations of the DNA sequence with directed evolution method. In this regard, both approaches, whether to bioprospect or to bioengineer biocatalysts have advantages and disadvantages which will be discussed in this paper.KEY WORDS: Sugar industry wastewater; aluminium sulphate; primary treatment, ferric chloride; polyaluminium chloride

KEYWORDS: *metagenomics, directed evolution, site-directed mutagenesis, error-prone PCR, DNA shuffling, high-throughput screening.*

1. INTRODUCTION

As natural catalysts, enzymes are regarded as one of the important classes of molecules to carry out a wide range of chemical reactions in our life. This is due to the key role of the enzymes in reducing the activation energy which accelerates the reaction reaching equilibrium [1]. However, the huge potential of enzymes is not limited to the essential reactions which usually occur inside the living cell. Biocatalysts are considered as sustainable and green tool in the new technological era including both the conventional applications such as food and detergent, as well as drug and bio-renewable energy of emerging industries [2]. High efficiency, stereoselectivity, chemo-selectivity, and regioselectivity, in an environment friendly conditions (mild temperature, pressure and pH) are the significant features of the catalysis offered by

enzymes. They drive enormous efforts in the fulfilment of industrial demands in terms of higher quality, new product and improved efficiency [3].

Despite the advantages of natural catalysts, they often encounter certain issues such as heat inactivation, instability against presence of denaturants, salinity, heavy metals, organic solvents and autolysis which cause the industrial exploitability more challenging [4]. Moreover, the unnatural aspect of the chemical reactions requires stable, functional and active catalysts that tolerate a wide range of conditions and generate considerable amount of product. These limitations have been also tackled by researchers' attempt for genetic diversity in order to discover new or engineer existing enzymes for novel properties. The past decades have seen a tremendous development of genetic diversity creation technologies [5].

There are several strategies to obtain functional enzymes with desired physico-chemical properties. Commonly, screening wild-type living organisms from the nature is regarded as one of the most effective approach. Prokaryotes and archaebacteria are generally the two most widespread groups of microorganisms known for providing highly diverse enzymes [6]. Beside all available methods, the recent advances in the technological research and innovations such as recombinant DNA technology and next-generation DNA sequencing have enabled these biocatalysts to meet the need of various industrial applications. Metagenomic studies, directed evolution and rational design are the available methods in order to evolve new enzymes. This necessitates construction of high-quality library for its success. Fig. 1 illustrates the three methods of enzymes finding.

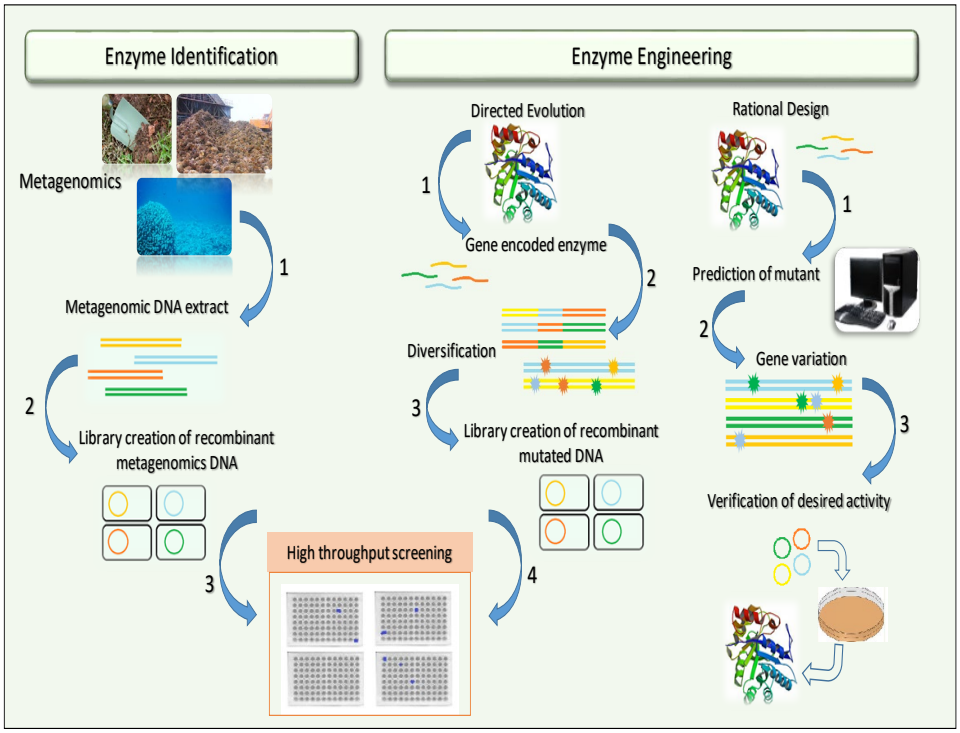

Fig. 1: Schematic illustration of enzymes finding by metagenomic approach, directed evolution and rational design.

2. FROM GENOMICS TO METAGENOMIS

It is well known that molecular microbiology misses up to 99% of microbial diversity by following the traditional microbiology techniques based on the bacterial cultivation. At this level, only 1% of microbial diversity is detected while the clear majority cannot be cultivated due to their strong adaptation to their niches [7-8]. By consequence, the possibility to bioprospect diverse enzymes is low. Since the apparition of great plate anomaly of the big and unexpected difference of microbe's number under microscopy and their number after cultivation, the notions of culture-dependent and culture-independent microbes appeared. This notion has triggered molecular microbiologists to explore another trend to solve this problem and to get access to the untapped bacterial reservoir. As the DNA is the responsible of microbial functions, the skip of cultivation step and obtaining of microbial DNA has solved the problem of unculturable microbes. The whole microbial DNA is called metagenome and the genetic study of this metagenome is called "metagenomic approach". The metagenomic approach is able to identify and classify a microbe. In addition, metagenomic is used to study metabolic activities and functional roles [9].

Due to the complexity of metagenomes, high-throughput screening methods are required to identify a desired gene. Metagenomic libraries can be screened by sequence-based or function-based method. The "DNA-sequence analysis" is to find out conserved regions of known gene to identify the similar families of the protein. In this approach, PCR or probe hybridization are the techniques frequently used [10]. The functional metagenomic approach requires no sequence information and it is based on the metabolic activity.

2.1 Sequence Metagenomic Approach

In this approach, it is possible to study the phylogenetic diversity of microbial population in a special habitat after the recovery of entire metagenomic DNA. This can be achieved by PCR amplification of 16S rRNAs using highly conserved primers of bacteria and archaea or by another marker gene like *dnaK* (HSP-70-type molecular chaperone) and *amoA* (ammonia monooxygenase) [11,12]. Bioinformatic tools are then used to analyze the data and phylogenetically classify the microbes regarding the taxonomic group. In addition, this approach allows to study microbial functions by comparing to the functions of known microbes in the databases or comparing to uncultured microbes discovered by previous metagenomic studies [7]. Metagenomic DNA from different sources were entirely sequenced to study the microbial functional richness [13,14].

The sequence-based screening approach does not require library preparation and gene expression is not dependent to the bacterial host. In contrast, it is indispensable in function-based screening to construct a DNA library and chose the suitable expression host and this makes sequence-based approach more advantageous at this level. The other advantage of this method is that similar screening strategies can be used for different targets. The major limitation of sequence based metagenomic is the ability to screen for well-known genes proteins only [10].

2.2 Functional Metagenomic Approach

Functional metagenomic approach has proven to be an excellent tool for high-throughput identification and characterization of biocatalysts from environmentally associated microbial communities. The advantage of this type of selection is the ability to discover novel biocatalytic mechanisms. Unknown-genes cannot be identified based on sequence information only as they have been screened following functional metagenomic approach. Their sequences are divergent from other genes sequence of same biocatalytic activity and they may represent totally new class of biocatalysts.

The basis of this method is the construction of metagenomic DNA libraries by cloning partially digested metagenomic DNA to a suitable vector and transformed to a surrogate host (generally *E. coli*) where the genes can be expressed. The metagenomic DNA libraries may be classified in term of the insert size into two groups, small- and large-insert libraries. The small-insert libraries are constructed using plasmids vectors with the insert capacity less than 15 kb to identify single or small operons of desired enzyme. The large-insert libraries use cosmids or fosmids for up to 40 kb of inserted DNA, or bacterial artificial chromosomes (BAC) vectors for more than 40 kb for the detection of large gene clusters. The use of BAC as vector is rare even though the advantage of very large-insert capacity, this is due to the difficulty of handling large DNA unit during cloning and screening. The high cost of such system is also a major inconvenient. To be able to analyze all the metagenome in small-insert libraries, large numbers of clones is needed. Instead, the convenient way is to screen small number of clones of large-insert libraries and cover all the metagenome. The screening of large-insert libraries is facing some problems even though it is considered as a powerful tool for isolation of new or improved biocatalysts. Weakly-expressed genes may not be detected due to low copy-number and some genes expression is limited by vector promoters [10].

The metagenomic DNA libraries can be screened in several ways. The first type of screening is when the library is screened phenotypically by the expression of a particular phenotype. In this assay the clone needs to express the protein to be screened extracellularly. The second type is the selection method where only clones with the interested activity can grow in certain conditions or metabolize some special substrates or resists some antimicrobial agents [15-17]. These methods are characterized by the low sensitivity and low throughput. In the last few years, the development of new functional metagenomic screening methods has enhanced the identification of novel biocatalysts missed by previous methods. SIGEX (substrate-induced gene expression) screening is one of the high-throughput screening methods developed by Uchiyama et al. [18] based on catabolic expression where a regulatory element is fused near to the gene that controls its induction by a specific substrate. Usually this method is combined with a fluorescence-based screening technology like FACS (fluorescent-activated cell sorting) and the *gfp* gene (gene encoding green fluorescent protein). PIGEX (product-induced gene expression) is also a high-throughput screening method based on the sensitivity of a special transcriptional activator to the product of target reaction [19]. METREX (metabolite-regulated expression) is a rapid screening where the metagenomic DNA clone contains the activity sensor. Compared to SIGEX, METREX identifies the metabolites activated by certain promoters, while SIGEX detects such promoters that are responsive to certain metabolites [20].

One of the major disadvantages of function-based screening is its dependency on expression of the cloned genes by the bacterial host. *E. coli*, which in some cases does not have the right chaperone needed for protein folding events after transcription and translation of the DNA. This problem can be circumvented by transferring the DNA after library construction in *E. coli* to another host such as *Ralstonia metallidurans* where the screening can be effectively conducted [21]. In Table 1, twenty-one biocatalysts identified by functional metagenomic approach are cited as examples. *E. coli* was frequently used as surrogate host for enzyme screening because of its availability in the market and it is well engineered for cloning and screening purposes. Other bacterial species or eukaryote hosts may be used for high-yield obtainment and to overcome the expression problems of eukaryote genes in prokaryotic system. Most researchers have adopted the agar and substrate screening method despite the development of high- and ultrahigh-throughput screening methods due to the high cost of the latter methods and the need of automated systems to handle a huge number of clones.

Finding the gene encoding the targeted biocatalyst is a crucial part in functional-metagenomic, and to get the full sequence of the gene, DNA sequencing needs to be carried out. Nowadays, next generation sequencing (NGS) is adopted in metagenomic approaches due to its capacity to merge huge data by covering big library size in one run and due to its high sensitivity and accuracy [22]. NGS strategy overcomes the problem of the high cost and time consuming of genome sequencing by Sanger sequencing approach; it allows the shearing of the genome into small fragments and re-assemble them after sequencing [23]. The NGS-data analysis may be a dilemma if the wrong analysis strategy is followed. The data size should be considered during data-analysis workflow, where the software used should be able to recover all the data. As Illumina sequencing strategy is based on the fragmentation of DNA to short reads before sequencing, an assembly step is compulsory to overlap the genome. The assembly method depends on the type of sequenced DNA, where genomic DNA or transcriptome cDNA is not assembled in the same manner. The known DNA is easily assembled based on homology alignment while the unknown DNA which has no reference template needs to be assembled with *de novo* assembly strategy [24].

2.3 Screening Enhancement by Enrichment Tool

Increasing the possibility of detecting a special gene can be done by favouring the growth of microbes containing it with one of enrichment strategies; by providing a growth substrate containing carbon or nitrogen as a source to select the microbes with the gene of interest [15,25]. As the enrichment step has the advantage of enhancement of gene detection frequency, it can also be a cause of decrease in metagenome diversity. Nevertheless, the combination of this strategy with functional metagenomic screening adopted by many researchers showed this combination to yield more positive clones during screening [26,27].

Table 1: Summary of some bioprospected enzymes through functional metagenomic approach from different habitats.

No	Enzyme	Metagenome's habitat	Host/Vector	Screening technique	Positive clones /number of clones	Ref
1	Alcohol oxidoreductase	Sugar beet field, river Grone, Solar lake, Gulf of Eilat	*E. coli*/Plasmid	Agar plate-based screening	24/400,000	[26]
2	Amidases	Agricultural field	*E. coli*/Plasmid	Agar plate-based screening	7/80,000	[28]
3	5 Esterases	Brine: seawater interface, Uranian hypersaline basin	*E. coli*/Phagemid	Agar plate-based screening	5/part of the library	[29]
4	13 β-lactamases	Remote Alaskan soil	*E. coli*/ Fosmid	Agar plate-based screening	14/337,000	[30]
5	11 Amidases	Activated sludge from aeration tank of a coke plant; wastewater treatment plant, Japan	*E. coli*/ benzoate-responsive sensor plasmid	Fluorescent substrate screening	143/96,000	[19]
6	Lipase	Qiongdongnan basin, South China Sea	*E. coli*/ Fosmid	Agar plate-based screening	19/200,000	[31]
7	Laccase	Mangrove soil	*E. coli*/ Plasmid	Chromogenic substrate screening	1/8,000	[32]
8	2 cellulases	Antarctic Peninsula	*E. coli*/Fosmid	Agar and substrates-based screening	2/52.000	[33]
9	2 Serine proteases	Surface sand from Gobi and Death Valley deserts	*E. coli*/Plasmid/Fosmid	Sequence-based screening	Not mentioned	[34]
10	Alkaline serine protease	Goat skin surface	*E. coli*/Plasmid	Agar plate-based screening	1/70,000	[35]
11	5 cellulases	offshore deep sub-surface petroleum reservoir 2.5 km below the ocean floor off the coast of Norway	*E. coli*/Fosmid	Agar and substrates-based screening	5/11,520	[36]
12	Xylanase	China Holstein cow rumen	*E. coli*/Fosmid	Agar plate-based screening	Not mentioned	[37]
13	β-glucosidases	Rumen from fistulated, non-lactating Holstein cows (Wales, UK)	*E. coli*/Fosmid	Agar and substrates-based screening	8/17,000	[38]
14	3 carboxylic ester hydrolases	Forest soil	*E. coli*/Plasmid	Chromogenic substrate screening	3/ not mentioned	[39]
15	Cold-adapted ß-galactosidase	Topsoil samples, Daqing oil field, Heilongjiang Province in China	*E. coli*/Plasmid	Agar and substrates-based screening	3/12,000	[40]
16	Phytases	Soil	*E. coli*/ Fosmid	Agar plate-based screening	2/14,440	[41]
17	Endocellulase and Endoxylanase	compost microbial communities that were collected from various farms (Thailand)	*E. coli*/Fosmid	Agar and substrates-based screening	1/ 1,500	[42]
18	β-Galactosidase	Ikaite columns (Greenland)	*E. coli*/BAC	Agar and substrates-based screening	2/2,843	[43]
19	6 Sulfatases and 8 Phospho-triesterases	different types of soil	*E. coli*/Plasmid	Picodroplets ultrahigh-throughput screening	6/1,250,000 8/1,250,000	[44]
20	Esterases	sixteen different environments	*E. coli*/Fosmid and lambda-ZAP	Agar and substrates-based screening	80/not mentioned	[45]
21	nitrile-metabolizing enzymes	10 soil samples collected from terrestrial and aquatic micro-environments in Waterford, Ireland.	*E. coli*/Fosmid	Agar and substrates-based screening	33/1.2 x 10^4	[46]

3. DIRECTED EVOLUTION

Over million years of evolution, organisms had found the way to adapt to the changes occurring around them by introducing mutations or polymorphisms in order to survive to the natural selection and the challenges that they face in the natural world. Throughout this evolution and adaptation, enzymes gained new features and traits regarding affinity, activity, selectivity, stability and solubility. However, the spontaneous mutational rate and evolution of a gene is generally low and not sufficient to generate a desired variant on a time scale [47], since the natural evolution is a process by which a pressure is employed to organisms during many generations and only those with high reproductive fitness survives.

Directed evolution is a method that mimics the natural evolution process in test tubes, taking benefit from the natural adaptation property of organisms towards evolution, in order to fulfil the industrial demands for biocatalysts that find its application in many industrial sectors, such as agriculture, medicine and health, veterinary and biotransformation of materials. The clue of a directed evolution framework is to generate a variants library by creating a genetic diversity to accelerate and enhance the prospect of the desired variant with the sought property, either using focused or random mutagenesis; genetic diversity can be achieved using various techniques and approaches including error-prone PCR (epPCR), megaprimer PCR of whole plasmid (MEGAWHOP), mutator strains, mutator plasmid and DNA shuffling. Most successful directed evolution researches combine both random and focused mutagenesis. Identification and isolation method are the second aspect of laboratory evolution. It allows the evaluation of every protein produced in the variant library for the desired property and selection of only the functional variants among the non-functional variants that contains the library. The common high throughput screening methods used are microtiter plates, fluorescent activated cell sorting (FACS), *in vitro* compartmentalization (IVTC), green fluorescent protein (GFP), digital imaging (DI), display (plasmid, phage display) and more [48]. Table 2 shows some examples of enzymes improved by directed evolution approach.

3.1 Error-Prone PCR (epPCR)

EpPCR is the most often used technique in directed evolution approach; it uses the mutation ability of the *Taq* Polymerase in order to randomly introduce mutations in the target gene sequence. *Taq* polymerase is used because of its natural high error rate under certain conditions. However, *Taq* polymerase produces a high level of biases of AT to GC which may not be suitable. Stratagene® has commercialized its own epPCR protocol using an engineered polymerase “Mutazyme” that has a more balanced mutation biases of AT to GC. The polymerase is commercialized in the “GeneMorph Random Mutagenesis Kit” with two releases I & II. Mutazyme II shows an enhanced mutation biases balance compared to *Taq* polymerase and Mutazyme I [49].

When using *Taq* polymerase, a typical epPCR protocol will include different variations of the PCR buffer by increasing the concentration of $MgCl_2$, adding $MnCl_2$ to the mixture, varying the ratios of dNTP, varying the template concentration, annealing temperature and the number of PCR cycles may also lead to the diminution of *Taq* polymerase fidelity giving then a number of mutations that were randomly inserted into the DNA sequence. The number of mutations or

mutation rate which can be defined as the number of mutations per kb depends on the ratios and concentrations used in the PCR protocol and also on the gene sequence and its length and GC % as well. A fraction of the epPCR products may be sequenced in order to determine the mutation rate of the epPCR protocol and the bias ratios, a matrix of the point mutations shall be drawn, and statistical estimation of the library can be calculated using some online computational programs [50].

3.2 DNA Shuffling

DNA shuffling was first developed and described by Stemmer in 1994 as an *in vitro* recombination method of homologous genes [51]. It is a powerful directed molecular evolution technique that results in a chimeric gene sequence that will lead to improved chimeric enzymes with better properties. DNA shuffling involves the fragmentation of homologous genes by DNaseI nuclease, followed by a recombination step which can be done through a number of denaturation, annealing, extension cycles by a polymerase where fragments anneal in a region of high sequence identity. A PCR is then applied in order to generate full-length chimeras using appropriate primers, and the PCR products will be cloned in an expression host to validate the shuffling by screening a library of chimeric enzymes produced by DNA shuffling of the parent gene. Several rounds of shuffling may be applied if the desired improvement is not achieved at the first recombination round. If a mutation is desired along with the recombination, epPCR can be used prior to the DNaseI digestion to generate random mutations in the parent genes that will be followed by recombination of the digests [52].

3.3 Mutator Strain

The preceding mutagenesis methods are *in vitro* methods using a simple thermocycler apparatus. Mutator strains is an *in vivo* mutagenesis method that rely on the ability of some bacterial strains that are basically lacking or deficient in the DNA proofreading and editing machinery that will lead to a high rate of spontaneous mutations throughout the recombinant plasmid. Mutator strains method has the advantage of the simplicity as no specialized equipment is needed, and basic knowledge of cloning technique is sufficient to create a large random mutant library with a significant mutation rate of the plasmid harbouring the gene of interest [2].

The common deficiencies used are the mutation of mutD, mutS and mutT genes with a predominance of the mutD gene deficiency preventing the exonuclease activity of the DNA polymerase III, resulting in a non-repair of the point mutations incorporated by DNA polymerase during replication. Despite the easiness of the Mutator strain random mutagenesis method, it has some disadvantages compared to the *in vitro* methods.Random mutations will occur in the whole plasmid, targeting the gene of interest may not be possible and may also present a large number of false positive clones in screening when mutations affect the promoter or the plasmid leading to an increase in protein production rather than mutation within the gene itself. Some mutations will lower the growth rate of the Mutator strain as they may occur in replication of necessary genes, and then rapidly, clones with low mutations number will overgrow and prevent subsequent mutagenesis since their plasmids will be isolated for screening. Thus, several rounds of plasmid isolation, re-transformation and mutagenesis maybe required to achieve a meaningful and functional mutation rate of the target gene. Several culture days may be required to achieve a high mutation frequency adding to the fact that growth must be monitored using a minimal

media (short doubling time compared to a rich media) to prevent the loss of the Mutator genotype while a PCR based random mutagenesis technique gives same frequency within a couple of hours [53].

Table 2: Summary of some engineered enzymes through directed evolution approach

No	Enzyme	Origin	Screening technique	Novel or improved characteristics	Reference
1	ß-glucosidase	*Agrobacterium sp.*	High-Throughput Screening	Enzyme is more efficient and with expanded repertoire of acceptable substrates	[55]
2	Citramalate synthase	*Methanococcus jannaschii*	High-Throughput Screening	- The specific activity was enhanced over a wide temperature range (30 to 70°C). -No sensitivity to feedback inhibition by isoleucine -Improved activity	[56]
3	*ß*-1,3-galactosyltransferase	*C. jejuni.*	Ultrahigh-throughput screening	variant with broader substrate tolerance than the wildtype enzyme and 300-fold higher activity	[57]
4	Ketol-acid reductoisomerase and alcohol dehydrogenase	*E. coli* for IlvC and *L. lactis* for AdhA	High-Throughput Screening	NADH-dependent pathway enables anaerobic isobutanol production at 100% theoretical yield and at higher titer and productivity than both the NADPH-dependent pathway and transhydrogenase over-expressing strain.	[58]
5	Ketol-acid reductoisomerase	*S. exigua, L. lactis, Shewanella sp, M. aeolicus* and *A. acidocaldarius*	High-Throughput Screening	-Construction of five ketol-acid reductoisomerase with reversed cofactor preference - two of them have NADH-dependent catalytic efficiencies greater than the wild-type enzymes with NADPH	[59]
6	TrpB-derived biocatalysts	*Pyrococcus furiosus* and *Thermotoga maritima*	High-throughput screening	-High level of biocatalyst expression in the host cell. -Enhanced solubility. -Ability to synthesize enantiopure Trp analogues substituted at the 4-, 5-, 6-, and 7-positions	[60]
7	Polysialyltransferases	*N. meningitidis*	Ultrahigh-throughput screening	increased enzymatic activity and improved stability compared to the wildtype enzyme	[61]

3.4 Mutator Plasmids

Mutator strains method has been successfully applied to directed evolution of genes giving an improved phenotype. However, they lack in term of genetic stability and the gene of interest shall be cloned into the Mutator strain first and sub-cloned into a stable strain for screening. Mutator plasmid is another *in vivo* random mutagenesis directed evolution technique that involves the transformation of a Mutator plasmid into a normal and stable strain transforming temporarily it into a Mutator strain allowing random mutagenesis of the gene of interest and increase the bacterial mutation rate. By the next round, the Mutator plasmid will be inactivated since it carries a temperature sensitive origin of replication (Ori) making the strain stable again but with the introduced mutations [54]. In this approach, the Mutator plasmid carries a mutant version of mutD gene that will bind to DNA polymerase III. Once transformed into the bacterial strain it will lead to a Mutator strain without involvement of mutagenesis of bacterial chromosome. Following this, a library shall be constructed and selection/screening of the mutant library for the desired evolution.

4. RATIONAL DESIGN

Rational design is basically making an educated guess about which amino acids in an enzyme to alter, and then making the changes using targeted or site-directed mutagenesis of the corresponding gene. The success rate of rational design approach relies on our knowledge about sequence and structure of protein of interest. Numerous previous successes have shown that when 3D structural information is available, protein design can be much more precise and accurate [62-64]. It is obvious that the knowledge of 3D structure of the target enzyme is a precondition and basis for structure-based design. However, it is also possible to model the 3D structure if there are solved structures of homologous proteins using 3D prediction software. The complexity of the structure/function correlation in enzymes has demonstrated to be the factor impeding the general application of rational design.

As the most relevant rational design applications is proteins thermostability improvement, the full information about protein catalytic mechanisms in the rational design case should be available and the link between structure and function and mechanisms of thermal stability need to be well-established as well [65–67]. As the catalytic and thermal stability mechanisms of proteins are in general weakly represented, rational design is facing a major obstacle. To overcome rational design difficulties, few methods have succeeded to enhance the thermal stability of proteins. Disulfide bonds strategy adopted by Voutilainen et al., 2008 has succeed to increase the unfolding temperature of the cellobiohydrolase enzyme with G4C/M70C/S290T mutation by 4 °C compared to the wild-type [68]. Protein surface charge strategy was adopted for the same aim by Kotzia and Labrou [69] to increase the half-life of L-asparaginase from 2.7 h to 159.7 h at 50 °C by the Asp133Val mutation. Xiao et al. [70] used the homologous comparison strategy to increase the half-life of mesophilic pectate lyase from *Xanthomonas camperstris* and effectively the mutant R236F half-life was 23-fold more than that of the wild type. Based on another strategy called protein unfolding free energy applied on α-amylase, [71] succeeded to increase the half-life of the enzyme with P477V mutation from 3.2 min to 5.1 min at 60 °C. B-factors strategy was also adopted by [72] on the lipase B isolated from *Candida antarctica* to increase its half-life by 4.5-fold at 50 °C by affecting the A162C-K308C mutations. Similar to the previous strategies, proline editing approach was also adopted to get the half-life

of the mutant 4R-FPhu-MscFv higher from 59 min of the wild-type to 240 min at 40 °C [73]. Other examples are listed in Table 3.

Table 3: Summary of some enzymes produced via site-directed mutagenesis.

No	Source	Mutation	Optimum Temperature (°C)	Remarks	Ref
1	*Trichoderma reesei*	A35V	67.7	Optimum temperature was observed at 67.7°C	[74]
2	*Clostridium phytofermentans*	(N144I/N291K/ E158V)	60	The best mutant had 92%, 36%, and 46% longer half-lives at 60°C on carboxymethyl cellulose, regenerated amorphous cellulose, and Avicel, respectively.	[75]
3	*Clostridium thermocellum*	S329G	85	Increase thermostability by 7°C, half-life at 85°C by 8.5 times.	[76]
4	*Clostridium thermocellum*	G283P	85	Increase of 14-fold at 85°C without loss of catalytic activity	[77]
5	*Streptomyces griseoaurantiacus*	SGUV30 and SGUV5	80	Type of mutation followed to create these two mutants is UV strain improvement	[78]
6	*Clostridium thermocellum,*	m1 =31313333; m2 =11312122; m3 =23232332; m4 =22222332; m5 =22122332.	70	Endoglucanase is engineered by SCHEMA, a structure-guided, site-directed protein recombination method, or by consensus-guided mutagenesis combined with random mutagenesis using error-prone PCR.	[79]

4.1 Rational Computational Design

The ability to produce biocatalyst from scratch enables scientists to build synthetic enzymes for a series of different chemical reactions. One of the advantages of a rational design approach is the increased probability of beneficial mutations and a significant reduction of the library size and thus less effort and time to be applied for the screening of the library. This is especially advantageous if no high-throughput assay system is available. There are many successful examples of rational design approach, among the recent was the successful increase in the thermostability of a cellulosomal endoglucanase, Cel8A, from *Clostridium thermocellum* by 14-fold at 85 °C without loss of catalytic activity [77]. In that report one single mutation (G283P) was sufficient to produce a thermostable variant; increasing half-life by 14-fold at 85 °C. Glycine to proline mutations have been described before to increase the thermostability, however only at carefully selected sites via rational design approach.

4.2 Molecular Docking

Molecular docking is widely used to predict protein-ligand complexes and to screen large libraries for molecules that will modulate the activity of a biological receptor. Molecular docking remains an important tool for structure-based screening to find new ligands and chemical probes, it has enriched hit-rates and often confirming the predicted geometries of the docked complex [80,81]. Intermolecular complex formed between two or more molecules can be predicted via

molecular docking. Most docking algorithms are able to generate a large number of possible structures. A range of algorithms have been developed to handle the docking problem [82]. The aim of molecular docking is to achieve an optimized conformation for both receptor and ligand, such that the overall free energy of the system is at its minimum. Nowadays, protein–ligand or protein–protein docking plays a vital role in predicting the orientation of the ligand when it is bound to a protein receptor or enzyme using shape and electrostatic interactions to quantify it [83].

Computational molecular docking can be achieved through two interrelated steps: first by sampling conformations of the ligand in the active site of the protein; then ranking these conformations via a scoring function [84]. In a simple rigid-body system, the ligand is searched in a six dimensional rotational or translational space to fit in the binding site, which may serve as a lead compound [85]. With six degrees of translational and rotational freedom as well as the conformational degrees of freedom of both the ligand and protein, there are a huge number of possible binding modes between two molecules. Unfortunately, it would be too expensive to computationally generate all the possible conformations. Various sampling algorithms have been developed and widely used in molecular docking software. Genetic algorithms (GA) [86] used to perform sampling is another class of well-known stochastic methods. Degrees of freedom of the ligand are encoded as binary strings called genes. These genes make up the 'chromosome' which actually represents the pose of the ligand. Mutation and crossover are two kinds of genetic operators in GA. Mutation makes random changes to the genes; crossover exchanges genes between two chromosomes. When the genetic operators affect the genes, the result is a new ligand structure. New structures will be assessed by scoring function, and the ones that survived (i.e., exceeded a threshold) can be used for the next generation.

Besides Coulombic interactions, van der Waals interactions, hydrogen bonds also play an important role. The sum of all these interactions is approximated by a docking score using a scoring function, which represents potentiality of binding.

4.3 Molecular Dynamics (MD) Simulation

In general, atoms of a molecule are moving and have interaction each other. The movement of atoms can be calculated and computed using Newton's law of motion principle. Usually, the motions of atoms are viewed and analysed in MD by allowing atoms and molecules to interact in predetermined period. The Newton's equations of motion are solved for a system of these interacting particles to determine the trajectories of molecules and atoms, where molecular mechanics force fields are employed to define the forces between each particle and the potential energy. This method was first established by [87] and his group successfully employed MD in the liquid water simulation in 1974 which was considered to be the first MD of a realistic system [88].

Nowadays, MD is widely applied in modelling of biomolecules as it makes the detailed study of complex large number of particles molecular system possible by using numerical methods. It has been almost 40 years since the first MD simulation of a macromolecule of biological interest was published [89]. The first molecule studied was bovine pancreatic trypsin inhibitor, BPTI. MD simulation are important tools for understanding the physical basis of the structure and function of biological macromolecules. The early view of proteins as relatively rigid structures

has been replaced by a dynamic model in which the internal motions and resulting conformational changes play an essential role in their function. Simulations can provide the ultimate detail concerning individual particle motion as a function of time. This information can be used to address specific questions about the properties of a model system, often more easily than experiments on the actual system. Although experiments play an essential role in validating the simulation methodology, comparison of simulation and experimental data serve to test the accuracy of the calculated results [90]. Currently there are many free and commercial software packages available for conducting molecular dynamics simulation, some of which are listed in the Table 4. The two MD program GROMACS [91] and NAMD [92] are widely used MD simulation packages. GROMACS aims at fast simulations while NAMD is available for parallel computing of large systems [93]. In terms of speed, GROMACS and NAMD are far ahead of other packages.

Table 4: Some available molecular dynamics packages

Software	Features	Reference
AMBER (Assisted Model Building with Energy Refinement)	Designed primarily for nucleic acids and proteins	[94]
GROMACS (GROningen MAchine for Chemical Simulations)	Designed primarily for proteins and lipids. High performance (very fast)	[91]
NAMD (NAnoscale Molecular Dynamics)	Designed for large biomolecular systems. Relatively fast and easily scalable (parallel MD).	[92]
CHARMM (Chemistry at HARvard Molecular Mechanics)	Designed for proteins, lipids and nucleic acids.	[95]
ACEMD (Acellera MD)	ACEMD is a production molecular dynamics software specially optimized to run on NVIDIA graphics processing units.	[96]
LAMMPS (Large-scale Atomic/ Molecular Massively Parallel Simulator)	Models an ensemble of particles in a liquid, solid, or gaseous state	[97]

5. CONCLUSION AND FUTURE PROSPECTS

The various available strategies to obtain genetic diversity of enzymes have been addressed in this review. Enzymes obtained through rational design or directed evolution present better stability, specificity and level of expression which is highly demanded by the industry. However, the natural pressure and selection that natural enzymes are subjected to, makes metagenomic a promising approach. Nevertheless, the limited access to high throughput screening facility combined with huge post screening data is the major challenge. Rational design is evidently information intensive, and a significant disadvantage to this approach is that the required structural/mechanistic data are available for only a tiny portion of interesting catalysts. On the other hand, directed evolution does not require any understanding of the association between protein structure and function. Instead, it lets the experiment reveal the best solution to the engineering problem by mimicking natural selection concept in the laboratory. It is understood that metagenomic is a highly interesting method to find novel and diverse enzymes. Yet, the continuous and increasing market demand for efficient biocatalysts raised the need of a new approach in order to fulfil the market requirements. In this way, it is interesting that metagenomic shall be followed by a semi-rational design where the novel enzymes discovered through metagenomic will be improved by combining rational and computational tools with directed evolution techniques to limit the region of mutagenesis and restrict the size of library to be screened for the targeted characteristic. notwithstanding, the choice of the approach to be used for the engineering of enzymes for biocatalysis application will be primarily influenced by the availability of funding and technical staff. Secondly, the output of the research project, whether one is aiming at improving an enzyme's activity, affinity, selectivity or at looking for novel enzymes and/or sources of enzymes.

ACKNOWLEDGEMENTS

The authors acknowledge financial support from MOE (FRGS13-086-0327 grant) and MOSTI (02-01-08-SF0294, Science fund) for metagenomic and directed evolution work, respectively.

REFERENCES

[1] Cao J, Li Y, Tu N, Lv Y, Chen Q, Dong H (2016). Novel enzyme/exfoliated bentonite nanohybrids as highly efficient and recyclable biocatalysts in hydrolytic reaction. J. Mol. Catal. B Enzym., 132: 41–46. https://doi.org/10.1016/j.molcatb.2016.06.016

[2] Wang P (2009). Multi-scale features in recent development of enzymic biocatalyst systems. Appl. Biochem. Biotechnol., 152:343–352. https://doi.org/10.1007/s12010-008-8243-y

[3] Yang G, Ding Y (2014). Recent advances in biocatalyst discovery, development and applications. Bioorganic Med. Chem., 22: 5604–5612. https://doi.org/10.1016/j.bmc.2014.06.033

[4] Bilal M, Asgher M, Parra-Saldivar R, Hu H, Wang W, Zhang X, Iqbal HMN (2017). Immobilized ligninolytic enzymes: An innovative and environmental responsive technology to tackle dye-based industrial pollutants – A review. Science of the Total Environment, 576: 646–659.

https://doi.org/10.1016/j.scitotenv.2016.10.137

[5] Tee KL, Wong TS (2013). Polishing the craft of genetic diversity creation in directed evolution. Biotechnology Advances, 31: 1707–1721. https://doi.org/10.1016/j.biotechadv.2013.08.021

[6] Heux S, Meynial-Salles I, O'Donohue MJ, Dumon C (2015). White biotechnology: State of the art strategies for the development of biocatalysts for biorefining. Biotechnology Advances, 33: 1653–1670. https://doi.org/10.1016/j.biotechadv.2015.08.004

[7] Coughlan LM, Cotter PD, Hill C, Alvarez-Ordóñez A (2015). Biotechnological applications of functional metagenomics in the food and pharmaceutical industries. Frontiers in Microbiology, 6:672. doi: 10.3389/fmicb.2015.00672

[8] Torsvik V, Goksoyr J, Daae FL (1990). High diversity in DNA of soil bacteria. Appl. Environ. Microbiol., 56: 782–787.

[9] Langille MGI, Zaneveld J, Copraso JG, McDonald D, Knights D, Reyes JA, Clemente JC, Burkepile DE, Thurber RLV, Knight R, Beiko RG, Huttenhower C (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nat. Biotechnol., 31: 814-/21. https://doi.org/10.1038/nbt.2676

[10] Rolf D (2005). The metagenomics of soil. Nat. Rev. Microbiol., 3: 470–478. https://doi.org/10.1038/nrmicro1160

[11] Webster G, Embley TM, Prosser JI (2002). Grassland management regimens reduce small-scale heterogeneity and species diversity of β-proteobacterial ammonia oxidizer populations. Appl. Environ. Microbiol., 68: 20–30. doi: 10.1128/AEM.68.1.20-30.2002

[12] Yap WH, Li X, Soong TW, Davies JE (2005). Genetic diversity of soil microorganisms assessed by analysis ofhsp70 (dnaK) sequences. J. Ind. Microbiol. Biotechnol., 17: 179–184. https://doi.org/10.1007/BF01574691

[13] Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, Wu D, Paulsen I, Nelson KE, Nelson W, Fouts DE, Levy S, Knap AH, Lomas MW, Nealson K, White O, Peterson J, Hoffman J, Parsons R, Baden-Tillson H, Pfannkoch C, Rogers YH, Smith HO (2004). Environmental Genome Shotgun Sequencing of the Sargasso Sea, Science., 304: 66–74. DOI: 10.1126/science.1093857

[14] Warnecke F, Luginbuhl P, Ivanova N, Ghassemian M, Richardson TH, Stege JT, Cayouette M, McHardy AC, Djordjevic G, Aboushadi N, Sorek R, Tringe SG, Podar M, Martin HG, Kunin V, Dalevi D, Madejska J, Kirton E, Platt D, Szeto E, Salamov A, Barry K, Mikhailova N, Kyrpides NC, Matson EG, Ottesen EA, Zhang X, Hernandez M, Murillo C, Acosta LG, Rigoutsos I, Tamayo G, Green BD, Chang C, Rubin EM, Mathur EJ, Robertson DE, Hugenholtz P, Leadbetter JR (2007). Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. Nature, 450: 560–565. https://doi.org/10.1038/nature06269

[15] Entcheva P, Liebl W, Johann A, Hartsch T, Streit WR (2001). Direct cloning from enrichment cultures, a reliable strategy for isolation of complete operons and genes from microbial consortia. Appl. Environ. Microbiol., 67: 89–99. doi: 10.1128/AEM.67.1.89-99.2001

[16] Donato JJ, Moe LA, Converse BJ, Smart KD, Berklein EC, McManus PS, Handelsman J (2010). Metagenomic analysis of apple orchard soil reveals antibiotic resistance genes encoding predicted bifunctional proteins. Appl. Environ. Microbiol., 76: 4396–4401. doi:10.1128/AEM.01763-09

[17] Staley C, Johnson D, Gould TJ, Wang P , Phillips J, Cotner JB, Sadowsky MJ (2015). Frequencies of heavy metal resistance are associated with land cover type in the Upper Mississippi River. Sci.

Total Environ., 511: 461–468. doi: 10.1016/j.scitotenv.2014.12.069

[18] Uchiyama T, Abe T, Ikemura T, Watanabe K (2005). Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes. Nat. Biotechnol., 23: 88–93. https://doi.org/10.1038/nbt1048

[19] Uchiyama T, Miyazaki K (2010). Product-Induced Gene Expression, a Product-Responsive Reporter Assay Used To Screen Metagenomic Libraries for Enzyme-Encoding Genes. Appl. Environ. Microbiol., 76: 7029–7035. doi: 10.1128/AEM.00464-10

[20] Handelsman J (2004). Metagenomics: Application of Genomics to Uncultured Microorganisms. Microbiol. Mol. Biol. Rev., 69: 195–195. doi: 10.1128/MMBR.68.4.669-685.2004

[21] Craig JW, Chang FY, Brady SF (2009). Natural products from environmental DNA hosted in Ralstonia metallidurans. ACS Chem. Biol., 4: 23–28. https://doi.org/10.1021/cb8002754

[22] Henson J, Tischler G, Ning Z (2012). Next-generation sequencing and large genome assemblies. Pharmacogenomics, 13: 901–915. doi: 10.2217/pgs.12.72

[23] Li Z, Chen Y, Mu D, Yuan J, Shi Y, Zhang H, Gan J, Li N, Hu X, Liu B, Yang B, Fan W (2012). Comparison of the two major classes of assembly algorithms: Overlap-layout-consensus and de-bruijn-graph. Brief. Funct. Genomics, 11:25–37. https://doi.org/10.1093/bfgp/elr035

[24] Benbelgacem FF, Mat Isa MN, Abdelrahim MAM, Tumian OA, Bellag A, Parman IA, Noorbatcha IA, Salleh H (2018). Next Generation Sequencing-Data Analysis for Cellulose- and Xylan-Degrading Enzymes from POME Metagenome. Sains Malaysiana, 47: 2951–2960. http://dx.doi.org/10.17576/jsm-2018-4712-03

[25] Voget S, Leggewie C, Uesbeck A, Raasch C, Jaeger KE, Streit WR (2003). Prospecting for Novel Biocatalysts in a Soil Metagenome. Appl. Environ. Microbiol., 69: 6235–6242. doi: 10.1128/AEM.69.10.6235-6242.2003

[26] Knietsch A, Waschkowitz T, Bowien S, Henne A, Daniel R (2003). Metagenomes of complex microbial consortia derived from different soils as sources for novel genes conferring formation of carbonyls from short-chain polyols on Escherichia coli. J. Mol. Microbiol. Biotechnol., 5: 46–56. doi:10.1159/000068724

[27] Ben Belgacem FF, Bellag OA, Ahmad AA, NoorbatchaIA, Salleh HM (2017). Palm oil mill effluent metagenome for cellulose-degrading enzymes. J. Basic Appl.Res., 3: 101-106.

[28] Gabor EM, De Vries EJ, Janssen DB (2004). Construction, characterization, and use of small-insert gene banks of DNA isolated from soil and enrichment cultures for the recovery of novel amidases. Environ. Microbiol., 6: 948–958. https://doi.org/10.1111/j.1462-2920.2004.00643.x

[29] Ferrer M, Golyshina OV, Chernikova TN, Khachane AN, Martins dos Santos VAP, Yakimov MM, Timmis KN, Golyshin PN (2005). Microbial Enzymes Mined from the Urania Deep-Sea Hypersaline Anoxic Basin. Chem. Biol., 12: 895–904. https://doi.org/10.1016/j.chembiol.2005.05.020

[30] Allen HK, Moe LA, Rodbumrer J, Gaarder A, Handelsman J (2009). Functional metagenomics reveals diverse β-lactamases in a remote Alaskan soil. ISME J., 3: 243–251. https://doi.org/10.1038/ismej.2008.86

[31] Hu Y, Fu C, Huang Y, Yin Y, Cheng G, Lei F, Lu N, Li J, Ashforth EJ, Zhang L, Zhu B (2010). Novel lipolytic genes from the microbial metagenomic library of the South China Sea marine sediment. FEMS Microbiol. Ecol., 72: 228–237.https://doi.org/10.1111/j.1574-6941.2010.00851.x

[32] Ye M, Li G, Liang WQ, Liu YH (2010). Molecular cloning and characterization of a novel metagenome-derived multicopper oxidase with alkaline laccase activity and highly soluble expression. Appl. Microbiol. Biotechnol., 87: 1023–1031. https://doi.org/10.1007/s00253-010-2507-5

[33] The Significance of the 260/230 Ratio in Determining Nucleic Acid Purity [https://www.researchgate.net/file.PostFileLoader.html?id=53181c86d4c1182b348b4597&assetKey=AS%3A272470362263560%401441973404725]

[34] Neveu J, Regeard C, Dubow MS (2011). Isolation and characterization of two serine proteases from metagenomic libraries of the Gobi and Death Valley deserts. Appl. Microbiol. Biotechnol., 91: 635–644. https://doi.org/10.1007/s00253-011-3256-9

[35] Pushpam P, Rajesh T, Gunasekaran P (2011). Identification and characterization of alkaline serine protease from goat skin surface metagenome. AMB Express, 1: 1-3. https://doi.org/10.1186/2191-0855-1-3

[36] Taupp M, Mewis K, Hallam SJ (2011). The art and design of functional metagenomic screens. Current Opinion in Biotechnology, 22: 465–472. https://doi.org/10.1016/j.copbio.2011.02.010

[37] Cheng F, Sheng J, Cai T, Jin J, Liu W, Lin Y, Du Y, Zhang M, Shen L (2012). A protease-insensitive feruloyl esterase from China Holstein cow rumen metagenomic library: Expression, characterization, and utilization in ferulic acid release from wheat straw. J. Agric. Food Chem., 60: 2546–2553. doi:10.1021/jf204556u

[38] Del Pozo MV, Fernandez-Arrojo L, Gil-Martinez J, Montesinos A, Chernikova TN, Nechitaylo TY, Waliszek A, Tortajada M, Rojas A, Huws HA, Golyshina OV, Newbold CJ, Polaina J, Ferrer M, Golyshin PN (2012). Microbial β-glucosidases from cow rumen metagenome enhance the saccharification of lignocellulose in combination with commercial cellulase cocktail. Biotechnol. Biofuels, 5. doi: 10.1186/1754-6834-5-73

[39] Biver S, Vandenbol M (2013). Characterization of three new carboxylic ester hydrolases isolated by functional screening of a forest soil metagenomic library. J. Ind. Microbiol. Biotechnol., 40: 191–200. doi: 10.1007/s10295-012-1217-7

[40] Wang K, Li G, Yu SQ, Zhang CT, Liu YH (2010). A novel metagenome-derived β-galactosidase: gene cloning, overexpression, purification and characterization. Appl. Microbiol. Biotechnol., 88: 155–165. https://doi.org/10.1007/s00253-010-2744-7

[41] Tan H, Mooij MJ, Barret M, Hegarty PM, Harrington C, Dobson ADW, O'Gara F (2014). Identification of novel phytase genes from an agricultural soil-derived metagenome. J. Microbiol. Biotechnol., 24: 113–118. doi: 10.4014/jmb.1307.07007

[42] Sae-Lee R, Boonmee A (2014). Newly derived GH43 gene from compost metagenome showing dual xylanase and cellulase activities. Folia Microbiol. (Praha)., 59: 409–417. doi: 10.1007/s12223-014-0313-7

[43] Vester J, Glaring M, Stougaard P (2014). Discovery of novel enzymes with industrial potential from a cold and alkaline environment by a combination of functional metagenomics and culturing. Microb. Cell Fact., 13: 72. doi: 10.1186/1475-2859-13-72

[44] Coughlan LM, Cotter PD, Hill C, Alvarez-Ordóñez A (2015). Biotechnological applications of functional metagenomics in the food and pharmaceutical industries. Frontiers in Microbiology, 6: 672. doi: 10.3389/fmicb.2015.00672

[45] Popovic A, Hai T, Tchigvintsev A, Hajighasemi M, Nocek B, Khusnutdinova AN, Brown G,

Glinos J, Flick R, Skarina T, Chernikova TN, Yim V, Bruls T, LePaslier D, Yakimov MM, Joachimiak A, Ferrer M, Golyshina OV, Savchenko A, Golyshin PN, Yakunin AF (2017). Activity screening of environmental metagenomic libraries reveals novel carboxylesterase families. Sci. Rep., 7. doi: 10.1038/srep44103

[46] Soares Bragança CS, Dooley-Cullinane TM, O'Reilly C, Coffey L (2017). Applying functional metagenomics to search for novel nitrile-hydrolyzing enzymes using environmental samples. Biomater. Tissue Technol., 1: 1-5. doi: 10.15761/BTT.100010

[47] Packer MS, Liu DR (2015). Methods for the directed evolution of proteins. Nature Reviews Genetics, 16: 379–394. https://doi.org/10.1038/nrg3927

[48] Xiao H, Bao Z, Zhao H (2015). High throughput screening and selection methods for directed enzyme evolution. Ind. Eng. Chem. Res., 54: 4011–4020. doi: 10.1021/ie503060a

[49] Cirino PC, Mayer ML, Umeno D (2003). Generating Mutant Libraries Using Error-Prone PCR. In Methods in molecular biology. Volume 231. Edited by Arnold FH, Georgiou G. New Jersey: Humana Press; pp. 3–10.

[50] Hanson-Manful P, Patrick WM (2013). Construction and analysis of randomized protein-encoding libraries using error-prone PCR. In Methods in molecular biology. Volume 996. 2nd Edition. Edited by Gerrard JA. New Jersey:Humana Press; pp. 251–267.

[51] Stemmer WP (1994). DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. Proc. Natl. Acad. Sci. U. S. A., 91: 10747–51. doi: 10.1073/pnas.91.22.10747

[52] Arnold FH, Georgiou G (2003) Directed Evolution Library Creation. Methods and Protocols. New Jersey, Humana Press.

[53] Nguyen AW, Daugherty PS (2013) Production of Randomly Mutated Plasmid Libraries Using Mutator Strains. In Directed Evolution Library Creation. Volume 231. New Jersey: Humana Press, pp. 39–44.

[54] Arnold FH, Georgiou G, Selifonova O, Schellenberger V (2013.)Evolution of Microorganisms Using Mutator Plasmids. In Directed Evolution Library Creation. Volume 231. New Jersey: Humana Press, pp. 45–52.

[55] Kim YW, Lee SS, Warren RAJ, Withers SG (2004). Directed Evolution of a Glycosynthase from Agrobacterium sp. Increases Its Catalytic Activity Dramatically and Expands Its Substrate Repertoire. J. Biol. Chem., 279: 42787–42793. doi: 10.1074/jbc.M406890200

[56] Atsumi S, Liao JC (2008). Directed Evolution of Methanococcus jannaschii Citramalate Synthase for Biosynthesis of 1-Propanol and 1-Butanol by Escherichia coli. Appl. Environ. Microbiol., 74: 7802–7808. doi: 10.1128/AEM.02046-08

[57] Yang G, Rich JR, Gilbert M, Wakarchuk WW, Feng Y, Withers SG (2010). Fluorescence Activated Cell Sorting as a General Ultra-High-Throughput Screening Method for Directed Evolution of Glycosyltransferases. J. Am. Chem. Soc., 132: 10570–10577. https://doi.org/10.1021/ja104167y

[58] Bastian S, Liu X, Meyerowitz JT, Snow CD, Chen MMY, Arnold FH (2011). Engineered ketol-acid reductoisomerase and alcohol dehydrogenase enable anaerobic 2-methylpropan-1-ol production at theoretical yield in Escherichia coli. Metab. Eng., 13: 345–352. https://doi.org/10.1016/j.ymben.2011.02.004

[59] Brinkmann-Chen S, Flock T, Cahn JKB, Snow CD, Brustad EM, McIntosh JA, Meinhold P, Zhang L, Arnold FH (2013). General approach to reversing ketol-acid reductoisomerase cofactor dependence from NADPH to NADH. Proc. Natl. Acad. Sci., 110: 10946–10951. https://doi.org/10.1073/pnas.1306073110

[60] Romney DK, Murciano-Calles J, Wehrmüller JE, Arnold FH (2017). Unlocking Reactivity of TrpB: A General Biocatalytic Platform for Synthesis of Tryptophan Analogues. J. Am. Chem. Soc., 139: 10769–10776. https://doi.org/10.1021/jacs.7b05007

[61] Janesch B, Baumann L, Mark A, Thompson N, Rahmani S, Sim L, Withers SG, Wakarchuk WW (2019). Directed evolution of bacterial polysialyltransferases. Glycobiology, 29: 588–598. https://doi.org/10.1093/glycob/cwz021

[62] Lutz S (2010). Beyond directed evolution-semi-rational protein engineering and design. Current Opinion in Biotechnology, 21: 734–743. doi: 10.1016/j.copbio.2010.08.011

[63] Wei J, Zhou Y, Xu T, Lu B (2010). Rational design of catechol-2, 3-dioxygenase for improving the enzyme characteristics. Appl. Biochem. Biotechnol., 162: 116–126 https://doi.org/10.1007/s12010-009-8720-y.

[64] Chen CY, Georgiev I, Anderson AC, Donald BR (2009). Computational structure-based redesign of enzyme activity. Proc. Natl. Acad. Sci., 106: 3764–3769. doi: 10.1073/pnas.0900266106

[65] Gao SJ, Wang JQ, Wu MC, Zhang HM, Yin X, Li JF (2013). Engineering hyperthermostability into a mesophilic family 11 xylanase from Aspergillus oryzae by in silico design of N-terminus substitution. Biotechnol. Bioeng., 110: 1028–1038. https://doi.org/10.1002/bit.24768

[66] Ruiz-Blanco YB, Marrero-Ponce Y, Paz W, García Y, Salgado J (2013). Global stability of protein folding from an empirical free energy function. J. Theor. Biol., 321: 44–53. https://doi.org/10.1016/j.jtbi.2012.12.023

[67] Socha RD, Tokuriki N (2013). Modulating protein stability - directed evolution strategies for improved protein function. FEBS J., 280: 5582–5595. https://doi.org/10.1111/febs.12354

[68] Voutilainen SP, Puranen T, Siika-Aho M, Lappalainen A, Alapuranen M, Kallio J, Hooman S, Viikari L, Vehmaanpera J, Koivula A (2008). Cloning, expression, and characterization of novel thermostable family 7 cellobiohydrolases. Biotechnol. Bioeng., 101: 515–528. https://doi.org/10.1002/bit.21940

[69] Kotzia GA, Labrou NE (2009). Engineering thermal stability of l-asparaginase by in vitro directed evolution. FEBS J., 276: 1750–1761. doi: 10.1111/j.1742-4658.2009.06910.x

[70] Xiao Z, Bergeron H, Grosse S, Beauchemin M, Garron ML, Shaya D, Sulea T, Cygler M, Lau PCK (2008). Improvement of the Thermostability and Activity of a Pectate Lyase by Single Amino Acid Substitutions, Using a Strategy Based on Melting-Temperature-Guided Sequence Alignment. Appl. Environ. Microbiol., 74: 1183–1189. doi:10.1128/AEM.02220-07

[71] Deng Z, Yang H, Li J, Shin HD, Du G, Liu L, Chen J (2014). Structure-based engineering of alkaline α-amylase from alkaliphilic Alkalimonas amylolytica for improved thermostability. Appl. Microbiol. Biotechnol., 98: 3997–4007. doi: 10.1007/s00253-013-5375-y

[72] Le QAT, Joo JC, Yoo YJ, Kim YH (2012). Development of thermostable Candida antarctica lipase B through novel in silico design of disulfide bridge. Biotechnol. Bioeng., 109: 867–876. https://doi.org/10.1002/bit.24371

[73] Edwardraja S, Sriram S, Govindan R, Budisa N, Lee SG (2011). Enhancing the thermal stability of

a single-chain Fv fragment by in vivo global fluorination of the proline residues. Mol. Biosyst., 7: 258–265. doi: 10.1039/C0MB00154F

[74] Sandgren M, Gualfetti PJ, Shaw A, Gross LS, Saldajeno M, Day AG, Alwyn Jones T, Mitchinson C (2003). Comparison of family 12 glycoside hydrolases and recruited substitutions important for thermal stability. Protein Sci., 12: 848–60. doi: 10.1110/ps.0237703

[75] Liu W, Zhang XZ, Zhang Z, Zhang YHP (2010). Engineering of Clostridium phytofermentans Endoglucanase Cel5A for improved thermostability. Appl. Environ. Microbiol., 76: 4914-4917. doi: 10.1128/AEM.00958-10

[76] Anbar M, Lamed R, Bayer EA (2010). Thermostability Enhancement of Clostridium thermocellum Cellulosomal Endoglucanase Cel8A by a Single Glycine Substitution. ChemCatChem, 2: 997–1003. https://doi.org/10.1002/cctc.201000112

[77] Anbar M, Gul O, Lamed R, Sezerman UO, Bayer EA (2012). Improved thermostability of Clostridium thermocellum endoglucanase Cel8A by using consensus-guided mutagenesis. Appl. Environ. Microbiol., 78: 3458–64. doi: 10.1128/AEM.07985-11

[78] Kumar AK (2015). UV mutagenesis treatment for improved production of endoglucanase and β-glucosidase from newly isolated thermotolerant actinomycetes, Streptomyces griseoaurantiacus. Bioresour. Bioprocess., 2: 22. DOI 10.1186/s40643-015-0052-x

[79] Moraïs S, Stern J, Kahn A, Galanopoulou AP, Yoav S, Shamshoum M, Smith MA, Hatzinikolaou DG, Arnold FH, Bayer EA (2016). Enhancement of cellulosome-mediated deconstruction of cellulose by improving enzyme thermostability. Biotechnol. Biofuels, 9: 164. doi: 10.1186/s13068-016-0577-z

[80] Coleman RG, Carchia M, Sterling T, Irwin JJ, Shoichet BK (2013). Ligand Pose and Orientational Sampling in Molecular Docking. PLOS ONE, 8: 1–19.https://doi.org/10.1371/journal.pone.0075992

[81] Chen H, Yao K, Nadas J, Bode AM, Malakhova M, Oi N, Li H, Lubet RA, Dong Z (2012). Prediction of Molecular Targets of Cancer Preventing Flavonoid Compounds Using Computational Methods. PLOS ONE, 7: 1–9. https://doi.org/10.1371/journal.pone.0038261

[82] Leach AR (2001) Molecular Modeling: Principles and applications. Second Edition. England: Pearson Education Limited.

[83] Pagadala NS, Syed K, Tuszynski J (2017). Software for molecular docking : a review. Biophysical Review, 9: 91-102. https://doi.org/10.1007/s12551-016-0247-1

[84] Meng XY, Zhang HX, Mezei M, Cui M (2011). Molecular Docking: A powerful approach for structure-based drug discovery. Current Computer-Aided Drug Design, 7: 146–157. doi:10.2174/157340911795677602

[85] Alberg DG, Schreiber SL (1993). Structure-based design of a cyclophilin– calcineurin bridging ligand. Science 262: 248–250. doi: 10.1126/science.8211144

[86] Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ (1999). Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. Journal of Computational Chemistry, 19: 1639–1662. https://doi.org/10.1002/(SICI)1096-987X(19981115)19:14<1639::AID-JCC10>3.0.CO;2-B

[87] Rahman A (1994). Correlations in the Motion of Atoms in Liquid Argon. Physical Review, 136:

A405–A411. 1994. https://doi.org/10.1103/PhysRev.136.A405

[88] Stillinger F, Rahman A (1974). Improved simulation of liquid water by molecular dynamics. The Journal of Chemical Physics, 60: 1545–1557.

[89] McCammon JA, Gelin BR, Karplus M (1977). Dynamics of folded proteins. Nature, 267: 585–590. https://doi.org/10.1038/267585a0

[90] Karplus M, McCammon JA (2002). Molecular dynamics simulations of biomolecules. Nature Structural Biology, 9: 646–52. https://doi.org/10.1038/nsb0902-646

[91] Lindahl E, Hess B (2001). GROMACS 3 . 0 : a package for molecular simulation and trajectory analysis, Journal of Molecular Modeling. 7: 306–317. https://doi.org/10.1007/s008940100045

[92] Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K (2005). Scalable molecular dynamics with NAMD. Journal of Computational Chemistry, 26: 1781–1802. https://doi.org/10.1002/jcc.20289

[93] Bhandarkar M, Bhatele A, Bohm, A, Brunner R, Buelens F, Chipot C, Dalke A, Dixit S, Fiorin G, Freddolino P, Grayson P, Gullingsrud J, Gursoy A, Hardy D, Harrison C, H´enin J, Humphrey W, Hurwitz D, Krawetz N, Kumar S, Kunzman D, Lee C, Mei C, Nelson M, Phillips J, Sarood O, Shinozaki A, Zheng G, Zhu G (2009). NAMD User's Guide, version 2.7b1. Theoretical Biophysics Group, University of Illinois and Backman Institue.

[94] Case DA, Pearlman JW, Caldwell TE, Cheatham J, Wang WS, Ross CL, Simmerling TA, Darden KM, Merz RV, Stanton AL, Cheng JJ, Vincent MC, Tsui V, Gohlke H, Radmer RJ, Duan Y, Pitera J, Massova I, Seibel GL, Singh UC (2002). AMBER 7. University of California, San Francisco., (c).

[95] Brooks BR, Brooks CL, Mackerell AD, Nilsson L, Petrella RJ, Roux B, Karplus M (2009). CHARMM: the biomolecular simulation program. Journal of Computational Chemistry, 30: 1545–614. https://doi.org/10.1002/jcc.21287

[96] Harvey MJ, De Fabritiis G (2009). An implementation of the smooth particle-mesh Ewald (PME) method on GPU hardware. J. Chem. Theory Comput., 5: 2371–2377. Doi:10.1021/ct900275y

[97] Plimpton S (1995). Fast Parallel Algorithms for Short – Range Molecular Dynamics. Journal of Computational Physics, 117: 1–19. https://doi.org/10.1006/jcph.1995.1039