Journal of Advances in Science and Technology
Vol. 13, Issue No. 2, March-2017, ISSN 2230-9659

An Analysis upon Various Strategies for Redesign and Direct Evolution of Enzyme Engineering

Sudha Vipin Sharma[1]* Dr. Sunita Hooda[2]

[1] Research Scholar, Mewar University, Rajasthan

[2] Associate Professor, Department of Chemistry, Acharya Narendra Dev College (University of Delhi) Govind Puri, Kalyani, New Delhi

Abstract – Enzymes are biocatalysts evolved in nature to achieve the speed and coordination of nearly all the chemical reactions that define cellular metabolism necessary to develop and maintain life. The application of biocatalysis is growing rapidly, since enzymes offer potential for many exciting applications in industry. The advent of whole genome sequencing projects enabled new approaches for biocatalyst development, based on specialised methods for enzyme heterologous expression and engineering. The engineering of enzymes with altered activity, specificity and stability, using sitedirected mutagenesis and directed evolution techniques are now well established.

The use of enzymes in industrial applications has been recognised for providing clean processes with minimal impact on the environment. Enzyme engineering is undergoing the most profound and exciting transformation in its history. It promises unprecedented expansion in the scope and applications of modified or improved enzymes with desired physical and catalytic properties. Two complementary strategies are currently available: rational redesign and directed evolution. Although both approaches have met with great success, each has limitations. In this article, the perspectives for these enzyme-engineering strategies are discussed briefly.

Directed evolution has become the preferred engineering approach to generate tailor-made enzymes. The method follows the design guidelines of nature: Darwinian selection of genetic variants. This review discusses the different stages of directed evolution experiments with the focus on developments in screening and selection procedures.

Enzymes are increasingly being used in an industrial setting as a cheap and environmentallyfriendly alternative to chemical catalysts. In order to produce the ideal biocatalyst, natural enzymes often require optimization to increase their catalytic efficiencies and specificities under a particular range of reaction conditions. A number of enzyme engineering strategies are currently employed to modify biocatalysts, improving their suitability for large-scale industrial applications. These include various directed evolution techniques, semi-rational design techniques, and more recently, the de novo design of novel enzymes.

----------♦----------

INTRODUCTION

Enzymes in their native form suffer several limitations for use as industrial catalysts, such as instability at high temperature, narrow substrate spectrum or insufficient enantioselectivity. These limitations may be overcome by engineering the enzyme at the genetic level or by modification of the enzyme without structural changes e.g., immobilization to a support. The former approach is commonly termed protein engineering, and involves introduction of changes in the gene coding for the enzyme, e.g., by rearrangement of the order of the amino acid sequence (circular permutations) or by introducing mutations either randomly or rationally. In general, protein engineering can be achieved via three main strategies: directed evolution, rational design and a hybrid of these two approaches known as semi-rational approach (Figure 1). The application of these tools has resulted in the successful engineering of biocatalysts from different enzyme families and has been reviewed elsewhere.

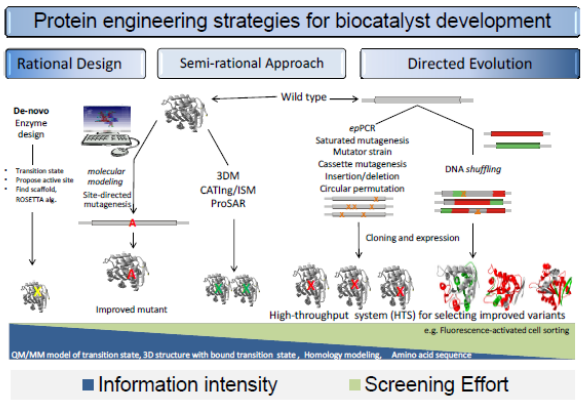

Figure 1 Illustration of the main strategies utilized for engineering proteins for the development of new biocatalysts. The acronyms are as follows: epPCR, error prone PCR; CASTing, Combinatorial Active Site Saturation Test; ISM, iterative saturation mutagenesis; ProSAR, Protein Sequence Activity Relationship.

In rational redesign, precise changes in amino acid sequence are preconceived based on a detailed knowledge of protein structure, function and mechanism, and are then introduced using sitedirected mutagenesis. This technology holds strong promise for optimizing the desired properties for commercial applications. It also greatly enhances our basic understanding of enzyme binding and catalytic mechanisms, thus increasing the success of future enzyme engineering efforts and laying the foundation for functional prediction of new protein sequences in databases. The power of rational redesign has been demonstrated by the generation of a faster superoxide dismutase, already one of the fastest known enzymes in nature and complete inversion of coenzyme specificities for both isocitrate and isopropylmalate dehydrogenases. In these studies, individual amino acid substitution or secondary structure engineering generated enzymes with desired properties. However, despite these spectacular examples, numerous attempts at redesigning enzymes have failed. These failures might result, to some extent, from an incomplete understanding of the underlying mechanisms required to enhance the desired enzyme properties. However, what probably accounts for many failed 'rational' engineering attempts is that a significant number were based on primary amino acid sequence homologies as the only criterion for amino acid replacements. In many cases, these substitutions were made without regard to the structural properties of the protein. Such 'homology-based engineering' frequently leads to substituting rigidly conserved amino acids that do not affect the desired enzyme properties and render the enzyme inactive because of changes in protein conformation. This process overlooks key amino acid residues, particularly when comparing highly divergent enzymes. The conventional approach of rational redesign requires confirmation of the mutation by sequencing and then by purification of the mutant enzymes following each round of mutagenesis in order that kinetic and functional properties can be determined. Such an approach is tedious and expensive, and might be impractical for multiple cycles of mutagenesis. Recently, an efficient strategy for identifying beneficial mutants using kinetics has been elaborated that would greatly facilitate the rational redesign of enzymes that require many cycles of mutagenesis to improve their properties.

Directed evolution does not require information about how enzyme structure relates to function. This technique employs a random process in which error-prone PCR is used to create a library of mutagenized genes. Genetic selection or highthroughput screening subsequently identifies the mutants that possess improved properties. The sorted genes might be subjected to further cycles of mutation and screening to enhance the original beneficial mutation. Recently, directed evolution has been significantly improved using in vitro recombination or DNA shuffling. These methods rapidly combine beneficial mutations that arise from random mutagenesis and significantly expand the sequence diversity derived from small pools of homologous genes. In the last few years, directed evolution has been widely adopted by industry and has proven extremely valuable for improving enzymes, as well as 'evolving' new metabolic pathways Although these techniques have been relatively successful in improving enzyme catalytic activity and physical properties, engineering substrate specificity appears to be more challenging. In the majority of cases, the enzymes had low catalytic activity and modest substrate specificity, and enzymes with new functions were rarely demonstrated. Evolutionary analysis of enzyme families suggests that drastic changes in enzyme function might require considerable changes in polypeptide backbones.

Such changes will probably not occur during the current in vitro evolution process, in which enzymes are mainly improved by point mutation with a significant bias for transitions over transversions (Table 1), thus limiting access to a broader spectrum of substitutions. In contrast, natural mutations ypically result from sexual or homologous recombination that generate deletions, insertions, duplications or fusions. Such mutations alter the spacing between amino acid residues and polypeptide chain segments and can result in large changes in specificity and new catalytic activities. Hence, a challenging task is to mimic the natural evolutionary process by introducing these natural mechanisms into directed evolution. Another limitation of directed evolution is the prerequisite for a sensitive and efficient method for screening a large number of potential mutants. Thus, development of novel enzyme assays suitable for high-throughput screening is needed to extend the applicability of

directed evolution to many more reactions of industrial interest.

	Natural evolution	Directed evolution
Point mutation	Moderate	Frequent
Deletion	Frequent	Rare
Insertion	Frequent	Rare
Inversion	Frequent	Rare
Duplication	Frequent	Rare
Fusion	Frequent	Rare
Recombination	Sexual and somatic	Sexual PCR
Selection	Natural selection (fitness)	Selection/high throughput screen

Table 1 : Comparison of mutational and selective mechanisms for enzyme improvement between natural and directed evolution.

	Rational redesign	Directed evolution
Knowledge of protein structure	Required	Not required
Knowledge of mechanism	Required	Not required
Point mutation bias	None	Transition favored
Secondary structure engineering	Feasible	Not feasible
Domain engineering	Feasible	Not feasible
Sensitive enzyme assay	Required	Not required
Selection scheme	Not required	Required

Table 2 : Comparison of enzyme engineering by rational redesign and directed Evolution.

Rational redesign and directed evolution both have their distinct advantages and yet the technologies are complementary. It has been clearly demonstrated that directed evolution can effectively augment rational redesign even though a great deal of knowledge about structure and function is available. In these studies, rational redesign was used to introduce key residues or structural elements that are not usually attainable via a random process. Random mutagenesis was then employed to generate subtle changes that would finetune protein packing and function. Conversely, random mutagenesis can also provide critical information for implementing a more 'rational' protein engineering strategy. The information can be used to minimize the sequence space that must be searched for future random mutagenesis experiments or to suggest targets for rational site-directed mutagenesis.

Enzymes are nature's biocatalysts catalyzing chemical reactions at high velocity, with great specificity, under mild temperatures, with water as solvent. These reaction conditions are regarded as energy efficient and environmentally friendly. The number of commercial enzyme applications is continuously growing, despite the suboptimal performance of many natural enzymes under industrial process conditions. The current limitations of applications of enzymes in industry are poor stability, low reaction rates, product inhibition, and limited substrate conversion. In addition, there are many reactions for which no enzymes are known today. Protein engineers are therefore focusing on the identification of enzymes with new reaction specificities and are improving the performance of existing enzymes. Also, DNA databases and environmental DNA libraries are screened for better performing enzymes, even though newly identified enzymes are likely to require additional optimization via protein engineering. Since the 1980s, site-directed mutagenesis (rational design) has been used to improve the properties of enzymes. Often, this approach has met with limited success, mainly because of a general lack of understanding of how protein structure relates to enzyme function. Nature, in contrast, applies Darwinian selection, for example, survival of the fittest, to alter the properties of enzymes. Since the 1990s, the Darwinian selection strategy has been applied in laboratory evolution of proteins. This approach, called directed evolution, has quickly proven to be much more effective in enzyme engineering than rational design. Directed evolution involves the generation of random genetic diversity followed by highthroughput screening for desirable variants (Fig. 2) and requires no structural knowledge of the protein. Where structural information is available, rational design and directed evolution are often combined to create ''smart libraries,'' introducing genetic variations at functional sites, such as the active site region of an enzyme.

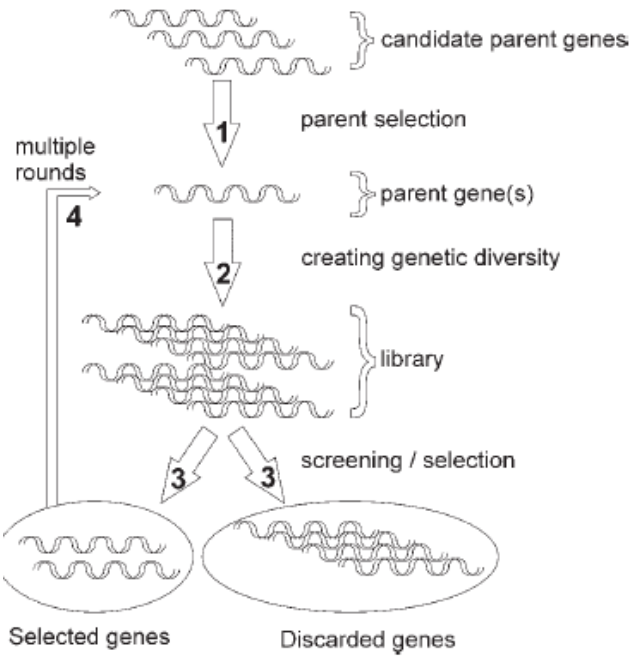

Figure 2. Flowchart of directed evolution.

ENZYME ENGINEERING

Another extremely promising area of enzyme technology is enzyme engineering. New enzyme structures may be designed and produced in order to

improve existing ones or create new activities. Over the past two decades, with the advent of protein engineering, molecular biotechnology has permitted not only the improvement of the properties of these isolated proteins, but also the construction of 'altered versions' of these 'naturally occurring' proteins with novel or 'tailor-made' properties (Gerlt and Babbitt 2009, Tracewell and Arnold 2009).

Affinity Tag	Matrix	Elution Condition
Poly-His	Ni^{2+}-NTA	Imidazole 20–250 mM or low pH
FLAG	Anti-FLAG monoclonal antibody	pH 3.0 or 2–5 mM EDTA
Strep-tag II	Strep-Tactin (modified streptavidin)	2.5 mM desthiobiotin
c-myc	Monoclonal antibody	Low pH
S	S-fragment of RNaseA	3 M guanidine thiocyanate, 0.2 M citrate pH 2, 3 M magnesium chloride
Calmodulin-binding peptide	Calmodulin	EGTA or EGTA with 1 M NaCl
Cellulose-binding domain	Cellulose	Guanidine HCl or urea > 4 M
Glutathione *S*-transferase	Glutathione	5–10 mM reduced glutathione

Table 3. Adsorbents and Elution Conditions of Affinity Tags.

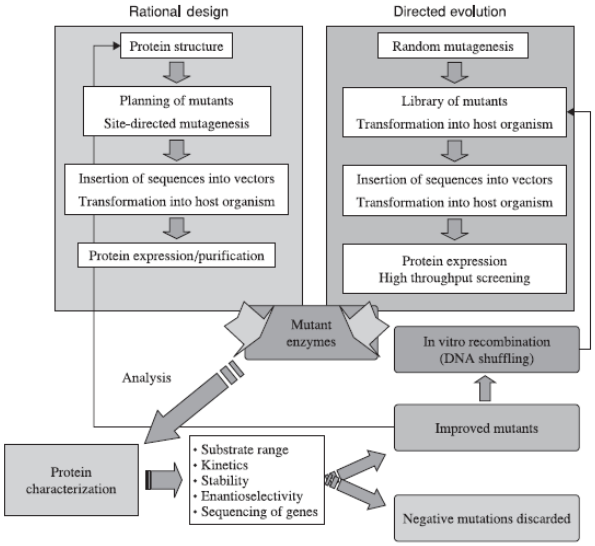

Figure 3. Comparison of rational design and directed evolution.

Tailor-Made Enzymes by Protein Engineering - There are two main intervention approaches for the construction of tailor-made enzymes: rational design and directed evolution (Schmidt et al. 2009).

Rational design takes advantage of knowledge of the three-dimensional structure of the enzyme, as well as structure/function and sequence information to predict, in a 'rational/logical' way, sites on the enzyme that when altered would endow the enzyme with the desired properties.Once the crucial amino acids are identified, site-directed mutagenesis is applied and the expressed mutants are screened for the desired properties. It is clear that protein engineering by rational design requires prior knowledge of the 'hot spots' on the enzyme. Directed evolution (or molecular evolution) does not require such prior sequence or three-dimensional structure knowledge, as it usually employs random-mutagenesis protocols to engineer enzymes that are subsequently screened for the desired properties . However, both approaches require efficient expression as well as sensitive detection systems for the protein of interest. During the selection process, the mutations that have a positive effect are selected and identified.

Usually, repeated rounds of mutagenesis are applied until enzymes with the desired properties are constructed. For example, it took four rounds of random mutagenesis and DNA shuffling of Drosophila melanogaster 2_-deoxynucleoside kinase, followed by FACS analysis, in order to yield an orthogonal ddT kinase with a 6-fold higher activity for the nucleoside analogue and a 20-fold kcat/Km preference for ddT over thymidine, an overall 10,000-fold change in substrate specificity.

The industrial applications of enzymes as biocatalysts are numerous. Recent advances in genetic engineering have made possible the construction of enzymes with enhanced or altered properties (change of enzyme/cofactor specificity and enantioselectivity, altered thermostability, increased activity) to satisfy the ever-increasing needs of the industry for more efficient catalysts (Zeng et al. 2009).

Rational Enzyme Design - The rational protein design approach is mainly used for the identification and evaluation of functionally important residues or sites in proteins. Although the protein sequence contains all the information required for protein folding and functions, today's state of technology does not allow for efficient protein design by simple knowledge of the amino acid sequence alone. For example, there are 10325 ways of rearranging amino acids in a 250-amino-acid-long protein, and prediction of the number of changes required to achieve a desired effect is an obstacle that initially appears impossible. For this reason, a successful rational design cycle requires substantial planning and could be repeated several times before the desired result is achieved.

DIRECTED EVOLUTION

Directed evolution is based upon the principle of natural evolution, whereby the incorporation of random mutations into the sequence of an enzyme allows the creation a large mutant library (10^3 - 10^6 mutants) displaying a high level of sequence diversity. This diversity is then explored by high-throughput screening to identify and select for those mutations which produce the desired phenotype or increase the enzyme activity, mimicking the process of natural selection. This selection procedure is repeated several times to produce the final biocatalyst with the desired traits.

The challenge of creating the expansive and diverse library of mutants called for by directed evolution has largely been overcome with the development of a number of robust techniques for producing genetic diversity. Perhaps the most commonly employed techniques to generate this diversity are error-prone PCR , which inserts mutations randomly across genes due to the fact that Taq polymerase lacks 3'–5' exonuclease proofreading activity; and DNA shuffling, which involves the recombination of homologous sets of genes. Other techniques to introduce sequence diversity include the use of mutator strains, which lack one or more DNA repair pathways; growth of cells harboring a plasmid encoding for the gene of interest in the presence of chemical mutagens such as EMS; and sequence saturation mutagenesis (SeSaM), which generates truly random mutations across each nucleotide within a given sequence .

One elegant example in which an enzyme was first selected for its novel activity, and the activity was subsequently catalytically improved by directed evolution, was presented by Seelig and Szostak. In order to aid their library screening process, they developed a technique known as mRNA display, which allows for the in vitro selection of enzymes from protein libraries. In this technique a DNA library is first created, and it is then transcribed into mRNA. A modified oligonucleotide containing puromycin (an antibiotic which resembles tRNA) is cross-linked to the 3' end of the mRNA before in vitro translation, resulting in mRNA-displayed protein. In order to carry out the selection process, the mRNAdisplayed protein is linked to the reaction substrate via reverse transcription of the mRNA to cDNA using a substrate-linked primer. Active enzymes can then be selected for as they will convert the substrate into the required product. The cDNA of the active enzymes is then isolated and is used for further rounds of evolution. Using this technique, Seelig and Szostak probed a library which had been prepared by mutating two recognition loops of the DNA binding domain of human retinoid-X-receptor using degenerate primers, preselection of random cassettes for intact open reading frames, and assembly of the final library by an iterative process of restriction and ligation. The authors tested the library consisting of 4×10^{12} RNA ligases for a particular novel activity – the ability to catalyze the ligation of a 5' triphosphorylated RNA oligonucleotide to the 3' hydroxyl group on a second RNA oligonucleotide. The activity of the resulting isolated RNA ligases was further improved by error-prone PCR. Following several rounds of mutagenesis and selection, 18 novel RNA ligases were found. The 7 most active ligases were expressed in E. coli as part of a maltose binding protein (MBP) fusion to improve stability and solubility of the proteins, and the most active of these fusion proteins was characterized. They found that their evolved RNA ligase was capable of catalyzing this novel reaction 2×10^6 times faster than the uncatalyzed reaction, which is a marked improvement.

One of the greatest advantages of the technique described by Seelig and Szostak, and in fact of directed evolution as a whole, is that no prior structural knowledge of the enzyme is required, permitting the engineering of enzymes whose function is not yet fully understood. However, the stochastic nature of directed evolution imposes a serious

limitation on this method – that is, the larger the library of mutants screened, the greater the chance of selecting the desired mutant. Consequently, this technique relies heavily on the ability to test the large number of mutants by a high-throughput assay, which is often an extremely labor-intensive process. The development of techniques such as mRNA display of proteins, fluorescence-activated cell-sorting (FACS) of cell-surface displayed mutants, and the incorporation of individual bacterial cells into microdroplets as a means of assessing gene expression and enzyme activity have made the screening of large mutant libraries a more practical and achievable process. Nevertheless, the creation of smaller, high-quality libraries containing more mutants displaying the required phenotype, as opposed to larger libraries consisting of a relatively high proportion of non-functional mutants, would be a more practical approach to circumventing the screening bottleneck. It is with this aim in mind that researchers have embarked upon the path of semi-rational design of biocatalysts.

RATIONAL DESIGN

Rational design in protein engineering science refers to the use of structural and mechanistic information to modify a protein. For the purpose of rational design, the three-dimensional (3D) structure should be known as it is essential in understanding at the molecular level the protein function and interactions with other protein or ligands (substrates). The protein structure is determined using either X-ray crystallography techniques or multidimensional NMR in the case of small proteins. Many enzyme structures have been solved at high resolution, which has paved the way for using computer modeling for protein engineering.

In addition, the advances of molecular modelling software, new generation of databases and computational tools allow non-experts to perform rational design for protein engineering. The approach of rational design was successfully applied in many cases to engineer enzyme selectivity and activity. Rational design has reached advanced levels with the

help of quantum mechanical calculations and information about the reaction mechanism. For example, Baker and co-workers have designed "de-novo" enzymes for reactions without known natural enzymatic catalyst.

Homology modeling - More than 6 million protein sequences are deposited in public databases nowadays, of which only 72 000 sequences have solved structures. This urges the computational biology scientists to develop algorithms to predict the 3D structure of proteins from their amino acid sequences. One effective approach is the prediction of the structure based on an already known structure (template) sharing similarity with the target sequence (40 % or higher). The process of homology modeling involves the following steps: secondary structure prediction, template identification, alignment of the target sequence over the template, model building, refinement and evaluation of the model. Bordoli et al. have described in detail the use of the SWISS-MODEL web server for obtaining a homology model for a protein sequence. In Paper I, the Yasara software was used for building a homology model of PLE-1 using mainly human carboxyl esterase (PDB 1MX9) as a template.

Molecular Modeling - Molecular modeling is an in-silico simulation of the behaviour of the molecules, which requires devising a number of equations and algorithms describing interand intra-molecular interactions. Application of quantum mechanics equations is highly accurate in describing the atom behaviour in molecules, especially of small molecules. However this approach is expensive in terms of computational processor units. Instead, simplified algorithms based on the molecular mechanics approach have been successfully used in a number of softwares. For example, AMBER, Assisted Model Building with Energy Refinement, is a family of force fields, i.e. a set of parameters and equations that describes the behaviour of atoms during the simulations, widely used for proteins and DNA simulations studies.

Docking - Docking refers to the prediction of the conformation that a small molecule (ligand or substrate) will acquire in the active site of a receptor macromolecule (enzyme, other types of protein or DNA), and estimation of the binding affinity. The docking technique was initiated for purposes of drug discovery i.e. insilico screening of a molecule library against certain receptors to find potential hits.

AutoDock4 is one of the most commonly used docking softwares. A more recent and improved version, AutoDock-Vina, has higher accuracy and shorter calculation time. Recently, AutoDock and AutoDock-Vina have been combined in a software package, PyRx, which allows easy running of the docking experiments.

CONCLUSION

Advances in the methodology available to the enzyme engineer have allowed the development of new strategies to design biocatalysts. While most early studies were focused on directed evolution techniques, there has been a shift towards employing directed evolution in conjunction with semi-rational design to produce biocatalysts. This has mainly been thanks to the introduction of a number of excellent computer algorithms, and to the exponential growth in the number of three dimensional structures and protein sequences which are available to the researcher. Smaller, high quality mutant libraries are now commonplace and a range of techniques have been introduced which permit the rapid and straightforward analysis of library members. We are now on the cusp of a new era of complete de novo design of biocatalysts, with successful reports of enzymes being designed to catalyze unnatural reactions already emerging.

Nowadays, directed evolution is the method of choice for enzyme engineering. Over the last few years, several selection and screening methods have been reported allowing screening of over 107 variants per day. The more traditional microtiter plate screenings are medium-throughput, but are compatible with most analytical tools. In an ever-evolving climate of laboratory enzyme evolution, it is difficult to predict which of the screening technologies will dominate in the future. However, to expand the scope of ultra-high-throughput screenings, these methods should move outside the limitations of fluorescence detection only. A promising development in this field is the recent report on alkaline phosphatase expressed from single cells in droplets using microfluidics, as this technology has the potential to integrate alternative analytical tools.

REFERENCES

Altamirano M.M. et al. (2000). Directed evolution of new catalytic activity using the alpha/betabarrel scaffold. Nature 403, pp. 617–622.

Bornscheuer, U. T. and Kazlauskas, R. J. (2004). Catalytic promiscuity in biocatalysis: using old enzymes to from new bonds and follow new pathways. Angew. Chem. Int. Ed. 43, pp. 6032–6040.

Chen, R. (1999). A general strategy for enzyme engineering. Trends Biotechnol. 17, pp. 344–345

Gerlt J. A., Babbitt P. C. (2009). Enzyme (re)design: lessons from natural evolution and

computation. Curr Opin Chem Biol 13(1): pp. 10–18.

Hult K. and P. Berglund (2007). Enzyme promiscuity: mechanism and applications. Trends Biotechnol., 25(5): pp. 231-8.

Matsumura I. (2000). IBC's Fifth Annual World Congress on Enzyme Technologies, Las Vegas

Reetz M. T. (2011). A method for rapid directed evolution. In: Lutz, S.; Bornscheuer, UT., editors. Protein Engineering Handbook. Wiley-VCH; 2011. pp. 409-440.

Reetz, M. T., Kahakeaw, D., and Lohmer, R. (2008). Addressing the numbers problem in directed evolution. Chembiochem. 9, pp. 1797–1804.

Schmidt M. et al. (2009). Protein engineering of carboxyl esterases by rational design and directed evolution. Protein Pept Lett 16(10): pp. 1162–1171.

Tracewell C. A., Arnold F. H. (2009). Directed enzyme evolution: climbing fitness peaks one amino acid at a time. Curr Opin Chem Biol 13(1): pp. 3–9.

Trinity-College-Dublin. ExplorEnz-The enzyme database. 2005; Available from: http://www.enzyme-database.org/index.php.

Valetti F, Gilardi G. (2004). Directed evolution of enzymes for product chemistry. Nat Prod Rep, 21: pp. 490-511.

Woycechowsky K. J., Hilvert D. (2004). Deciphering enzymes – genetic selection as a probe of structure and mechanism. Eur J Biochem, 271: pp. 1630-1637.

Zeng Q. K. et. al. (2009). Reversal of coenzyme specificity and improvement of catalytic efficiency of Pichia stipitis xylose reductase by rational site-directed mutagenesis. Biotechnol Lett 31(7): pp. 1025–1029.

Corresponding Author

Sudha Vipin Sharma*

Research Scholar, Mewar University, Rajasthan

sudha.sharma20@gmail.com