REVIEW

CrossMark

In-Silico Bioprospecting: Finding Better Enzymes

Asmita Kamble[1] · Sumana Srinivasan[2] · Harinder Singh[1]

Abstract

Enzymes are essential biological macromolecules, which catalyse chemical reactions and have impacted the human civilization tremendously. The importance of enzymes as biocatalyst was realized more than a century ago by eminent scientists like Kuhne, Buchner, Payen, Sumner, and the last three decades has seen exponential growth in enzyme industry, mainly due to the revolution in tools and techniques in molecular biology, biochemistry and production. This has resulted in high demand of enzymes in various applications like food, agriculture, chemicals, pharmaceuticals, cosmetics, environment and research sector. The cut-throat competition also pushes the enzyme industry to constantly discover newer and better enzymes regularly. The conventional methods to discover enzymes are generally costly, time consuming and have low success rate. Exploring the exponentially growing biological databases with the help of various computational tools can increase the discovering process, with less resource consumption and higher success rate. Present review discusses this approach, known as in-silico bioprospecting, which broadly involves computational searching of gene/protein databases to find novel enzymes.

Keywords In-silico · Bioprospecting · Enzyme

Introduction

Enzymes play an important role in our daily lives and are used in variety of industries and sectors like food, detergent and medicine [1]. The demand of certain enzymes has increased exponentially, like lipases, proteases, hydrolases and polymerases. Research laboratories and industries are extensively working to find newer and better candidates. Major enzyme industries are regularly introducing new enzymes in the market. In the past two decades, several patents on enzymes have been filed and issued. Apart from this, there are ongoing efforts to substitute chemical reaction processes in industries with enzymatic processes, as they are greener and environment friendly alternatives. It has been widely accepted that a cleaner chemical synthesis process should be practiced to prevent pollution and avoid generation of toxic wastes [2]. Enzymatic synthesis of chemical compounds has emerged as a simple, better and competitive route in comparison to chemical methods. Also, a high substrate specificity and better conversion rate with formation of low or no by-products makes enzyme a robust and efficient choice. Recently, Merck and Codexis developed a greener process for the synthesis of Sitagliptin, a drug used in diabetes treatment [3]. In the recent years, advancement in recombinant DNA technology has resulted in successful approaches to overexpress an enzyme in variety of host cells, which can help in producing the biocatalyst in high amount. To obtain an efficient enzyme candidate, stringent selection criteria are required to achieve high activity, specificity, and stability. In an industrial processes, the substrate, solvent, reaction conditions are important and an enzyme chosen should be able to withstand these components and conditions. It is actually difficult to find a natural enzyme with all the properties desired in an industrial process. To fulfil the massive enzyme demand, various approaches are practiced to constantly explore different resources to obtain new and better enzymes. Among these, in-silico bioprospecting has come up as an efficient, cost and time effective approach to discover new enzyme candidates. Although this approach has been practiced at various laboratories [4–6], it has not been reviewed or discussed.

✉ Harinder Singh
   Harinder.Singh@nmims.edu

1   Department of Biological Sciences, Sunandan Divatia
    School of Science, NMIMS Deemed to be University, Vile
    Parle (W), Mumbai 400056, India

2   Biosystems Engineering Lab, Department of Chemical
    Engineering, IIT Bombay, Powai, Mumbai 400076, India

In-Silico Bioprospecting

New enzyme discovery can be accomplished using various conventional and contemporary methods as mentioned in Fig. 1. Common methods of screening to identify novel enzymes are performed by exploring natural sources like industrial waste or soil, but they require an established protocol for screening assay or selection method based on the desired properties of the enzyme. This process involves biochemical screening and isolating the organism on selective media, which is usually time and resource consuming and may or may not result in a novel candidate. From these screening assays, the selected organism further needs to be identified, followed by the identification of gene sequence which is coding for the desired enzyme and function. One approach is to perform random mutagenesis to create enzyme mutant, and then sequence the DNA region. Another way is to perform targeted or whole genome sequencing to identify the desired enzyme gene sequence. As an alternative, amplification of target gene can be performed using degenerate primers [7]. There are challenges involved in primer designing, which affects the success rate. The process is followed by PCR library cloning and screening for prospective candidates with desired properties, which again demands a well-established protocol for screening positive candidates. After selecting the desired clone, the responsible gene can be sequenced, cloned and expressed.

The direct screening and identification methods are preferred where molecular biology resources are inadequate. These experimental approaches are used commonly, but they are time and resource consuming, with low success rate. However, in-silico bioprospecting is a simple, straightforward and promising approach to identify novel enzyme candidates with better enzymatic properties. A compilation of recent reports, where in-silico bioprospecting approach has been used to find novel enzymes, is given in Table 1. The current fast paced, high-throughput whole genome/metagenome sequencing has tremendously increased the biological database and thus the enzyme diversity. This diversity in turn has increased the complexity and difficulty of finding a novel candidate. The in-silico bioprospecting process can be broadly divided into two steps: (i) Searching databases (ii) Using Bioinformatics tools to screen, analyse and shortlist prospective candidates.

Step 1: Searching Databases

This can be performed by exploring databases using various search tools based on homology, conserved motif, consensus guided approach, or simply keyword search. The search result can be further screened using filters, such as percentage identity, query coverage, *e*-value. For example, a keyword search in NCBI protein database can be performed, followed by filtering the results to show candidates between 30 and 80% identity with query coverage > 95%. Gupta et al. [11] used keywords such as 'Hypothetical Protein of
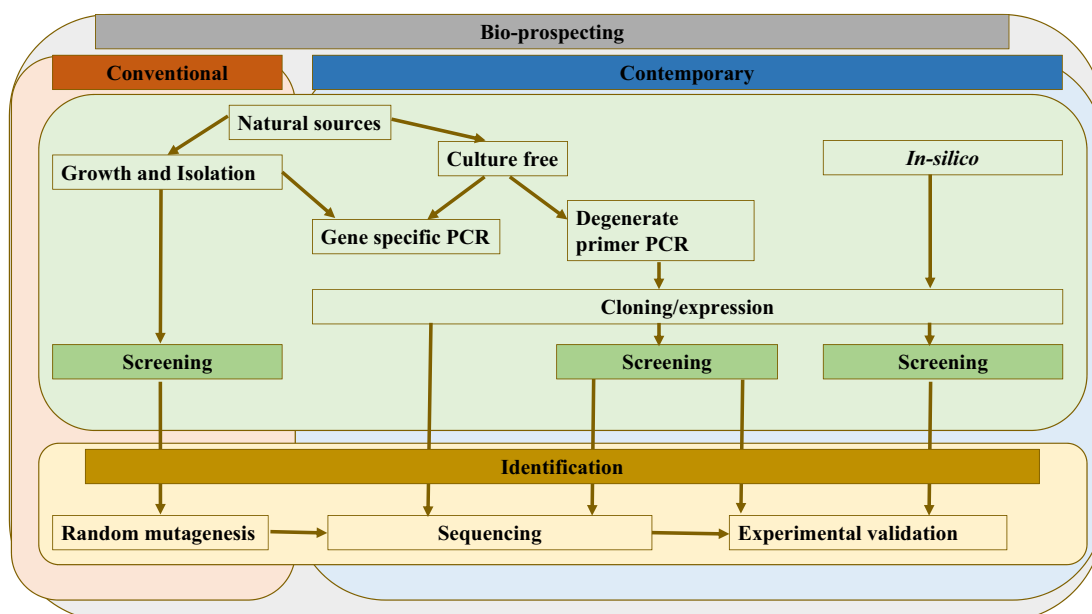

Fig. 1 Methods of enzyme bioprospecting

Table 1 In-silico bioprospecting approach used to find novel enzymes

Sr. no.	Enzyme	Approach/method	Reference
1	Histidine acid phytase (HAP)	Pfam identifier for the conserved domain Pfam00328 was used screen novel HAP metagenomes of the acidic peatland soil in Schlöppnerbrunnen, in the Fichtelgebirge Mountains, Germany Sequences were further manually analysed to examine the presence of the conserved motifs (RHGXRXP and HD) which were present in HAP phytase Further Phylogenetic analysis and taxonomic affiliations given by using MAFFT online service, nBLAST, PhyloPythiaS online tool Followed by Comparative homology modelling by using SWISS-MODEL Experimental validation	[8]
2	Aldehyde-deformylating oxygenase (ADO)	*Synechococcus elongatus* PCC7942 used as the query sequence to screen for the presence of *ado* gene in the set of sequences of organisms belonging to hot spring IMG/MER database Structural and Functional analysis performed by using various Bioinformatics tools such as ExPASy translate tool for translating sequence, MEGA software for phylogenetic analysis Adopted structure based protein engineering approach to improve thermostability Experimental validation	[9]
3	Nitrilases	Homology and motif based approach adopted for Genome screening Insilico analysis of putative sequences Experimental validation	[10]
4	Baeyer–Villiger and CYP153Monooxygenases	Metagenomic dataset created by isolating DNA sample from 23 sediment samples from (i) Advent Fjord, Spitsbergen, Svalbard Archipelago, Norway [NOR]; (ii) Port Värtahamnen, Stockholm, Baltic Sea, Sweden [SWE]; (iii) Ushuaia Bay, Tierra del Fuego Island, Argentina [ARG]; and (iv) Potter Cove, 25 de Mayo (King George) Island, Antarctica [ANT] was further submitted to IMG database Pfam domains (PF00067 for Cytochrome P450 or PF00743 for Flavin-binding monooxygenase-like) was used to screen this metagenomic dataset Further Blastp analysis performed using well characterized enzyme and/or crystallized sequences as reference Insilico characterization of putative sequences	[5]
5	Hypothetical protein	Hypothetical Protein was searched using NCBI database using keywords 'Hypothetical Protein of *Triticum aestivum*' or 'Hypothetical Proteins of wheat' Functional annotation of identified novel abiotic stress proteins of *Triticum aestivum*	[11]
6	Cellulose-and/or hemicellulose-degrading enzymes	Functional metagenomics Porcupine Microbiome metagenomic data created Relevant protein sequences of interest were selected on the basis of domain conservation, low *e*-values- pHMMER & the Research Collaboratory for Structural Bioinformatics Protein Data Bank Experimental validation of putative candidates	[12]
7	β-Glucosidase	Metagenomic DNA library constructed Search in the existing databases Experimental validation	[13]
8	b-(1,3) Galactosyltransferases	20 putative b-(1,3)-GalT genes identified by performing tblastn using three Homo sapiens b-(1,3)-GalT sequences: GalT1 (Q9Y5Z6), GalT2 (O43825), GalT4 (O96024) against *Arabidopsis thaliana* sequence database at NCBI Insilico characterization and experimental validation performed	[14]
9	Polyhydroxyalkanoate (PHA) synthase	Complete or nearly full-length PHA synthase genes retrieved from MG-RAST database mangrove soil metagenomic data Experimental Validation of novel and wide substrate specific PHA	[15]

T. aestivum', 'Hypothetical Proteins of wheat' in NCBI database followed by manual screening to get unique protein candidates. After removing redundant entries, unique candidates were further subjected to physicochemical, localization, function and domain analysis. In another database search, keywords such hydroxybutyrate, hydroxyalkanoate, hydroxyalkanoic, PHA and PHB were used as input [15]. Another common approach practiced by researcher is to search biological databases using a known candidate enzyme sequence. While choosing a potential enzyme gene sequence, it is of utmost importance to select a full length protein sequence having conserved domains, as many incomplete sequences annotated in database do not code for a functional protein, when checked experimentally. Also, in the search result, the selected candidate's sequence similarity should not be very high with known sequence. This is to ensure that a novel candidate is shortlisted and not a close homologue of a known sequence. In the similarity search result, the hits with > 90 identity are very closely related, sources like different species of same family, and it is more likely that they are very similar. But, the hits with ∼ 80% identity or lower are those candidates who are different from the query candidates, not closely related, but do have conserved sequences similar to known candidates. This ensures that novel candidates are chosen, which is predicted to retain the enzyme activity but is different from the search query. There have been reports where researchers had selected candidates with sequence similarity as low as 40 percent. Sharma et al. [10] searched novel sources of nitrilases from microbial genomes by adopting homology-based approach and selected sequences which exhibited > 30% and < 80% identity. The shortlisted search results need to be confirmed for a complete coding sequence or sequences. For example, shortlisted candidates of nitrilase were checked by GenMark S tool to verify complete coding sequences or sequences [10]. Since the protein length information is available for the input sequence, the search results should be restricted to length closer to the input sequence length. In case of nitrilases, sequences with less than 100 amino acids were considered as false positive and were discarded [10]. In another instance, sequences less than 250 amino acids were excluded to find novel BVMO (Bayer-Villiger Monooxygenases) enzyme [5]. For PHA synthase, sequences with ∼ 120 to 260 bp were considered as prospective candidate in a database search [15]. These search filters along with others like *e*-value, can aid in gathering positive sequences which could code for functional enzyme of appropriate length and reduces the chance of false discovery or random or irrelevant search result.

In certain cases, designing motif from selected protein sequences [*e.g.* by using MAST (Motif Alignment and Search Tool) at MEME suite] can be used to search bacterial genome. For example, Homology-based approach and motif

search resulted in the identification of 138 putative/hypothetical protein sequences which had potential to code for nitrilase [10]. Vaquero et al. [16] also adopted homology-based strategy to screen for novel CalB-type lipase in fungal genomes using blastp algorithm, against JGI and NCBI databases, with *e*-value cut-off as 10^{-2}. In the same study, conserved motif approach failed to identify putative lipase gene due to absence of conserved sequence motif generated by MEME software. Therefore, different individual strategies or combinations should be implemented in the process of finding novel putative enzymes. Consensus-guided approach, using Pfam domain, can also be used to search databases for the presence of particular enzyme family. Consensus-guided approach was adopted by Shakeel et al. [9] to obtain heat stable alkane-producing enzymes, using *ado* gene from *Synechococcus elongatus* PCC7942 as a query to search IMG/MER hot spring database. A consensus sequence was generated from the list of homologous sequences using Bioinformatics tools, which was further validated computationally and experimentally.

Specific datasets like metagenomes from various ecosystems can also be searched for obtaining novel enzymes. Around 264 putative monooxygenases were obtained when Pfam domain and blastp search were used to search BVMO [5] from ∼ 14 million protein-coding sequences present in metagenomic dataset of cold marine sediments [5]. Metagenome data of mangrove soil were explored to find polyhydroxyalkanoate (PHA) synthase genes [15]. Adam et al. [17] reported a novel activity-based approach to screen H_2-uptake enzyme from hydrothermal Metagenome. Toyama et al. [13] reported a novel β-glucosidase from microbial Metagenome of a lake in Amazon. Tan et al. [6] reported a novel thermostable phytase using bioinformatics approach which was screened from Metagenome database. Various steps and approaches used in gene mining from Metagenome data have been discussed and reviewed recently and reader is referred to these articles and reviews [18, 19] for details.

The steps of in-silico bioprospecting can be modified as per the desired property of enzyme. For example, if a thermostable enzyme is desirable, but the known enzyme reported is not thermostable, the similarity searches in thermophiles will be useful to find putative thermostable enzymes. It has been commonly observed that the thermostable enzyme sequences are different from their mesophilic counterpart. The putative thermophilic candidates searched this way should be further analysed (discussed in Step 2) to make sure that residues important for structure and functions are conserved.

Step 2: Using Bioinformatics Tools to Screen, Analyse and Shortlist Prospective Candidates

Once the primary list has been generated using various database search approaches, the next step will be to

analyse their physiochemical, phylogenetic and functional properties using different bioinformatics tools. ProtParam software using ExPASy server is widely used to access physiochemical properties (such as the molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index, and grand average of hydropathicity (GRAVY) of putative candidates [10, 11, 20]). Predicted values of all parameters of putative enzyme(s) are compared to the well characterized enzyme which affects the confidence level to study the putative enzyme(s) experimentally. For example, ProtParam predicted physiochemical properties of 138 putative nitrilases with in the range of well-characterized nitrilases [10]. All the parameters are based on protein sequence *i.e.* sequence-dependent analysis; therefore, it is necessary to get complete or nearly complete sequence for accurate analysis and prediction of various physiochemical properties.

Phylogenetic analysis can be performed using tools like Molecular Evolutionary Genetics Analysis (MEGA) [11, 15, 16, 21]. For example, phylogenetic analysis of selected putative candidates belonging to CalB-family grouped putative lipases in to different clusters of known lipases depending upon its evolutionary closeness [16], thus helping in deciding on novel and unique candidates. Structural modelling of putative candidates can be performed using SWISS-MODEL server or MODELLER v9.15 software [21]. Vaquoro et al. [16] used CalB as template to model PlicB, which exhibits 30% sequence identity and 44% similarity. The information about structure and residue conservation prediction is only possible if structural data of protein homologues are available through crystal structures. Hence, persistent exploration and enrichment of databases are necessary for in-silico bioprospecting of novel enzymes.

There are other tools which can predict structural information such as signal peptide (e.g. Signal P) or disulphide linkages (*e.g.* DiANNA). DiANNA 1.1 web server predicted two disulfide bonds in PlicB whereas CalB and Uml2 lacks disulfide bonds [16]. Protein functional domains and families are studied by comparing list of putative enzyme(s) against databases like Pfam, CATH, SVM-Prot, CDART, SMART. In one study, hypothetical proteins (HPs) were explored using tools based on domain architecture and profiles [11]. Out of 124 HPs, 77 sequences were annotated with high confidence by using Pfam, CATH, SVM-Prot, CDART, SMART and ProtoNet, and among them, 16 were predicted as enzymes. Functional protein network provides information about the association of hypothetical/putative protein(s) with the known functional protein, which can be generated by STRING database. In the study conducted by Gupta et al. [11], it was found that the predicted HPs such as HAV22 (Q7XAP6) and F-box protein (D0QEJ9) were interacting with other proteins of the STRING database such as protein 4,345,793 of *Oryza sativa* subsp. *Japonica.*

Analysing the putative candidates using bioinformatic tools provides clarity and help in selecting those candidates which are structurally and functionally more suitable, novel and unique. Following the sequence selection, candidates are validated for desired properties by cloning and expressing them in artificial expression systems followed by physiochemical characterization of enzyme [6, 13]. Apart from in-silico bioprospecting, enzymes with desired properties such as high activity [22, 23], substrate specificity [24] and stability [25, 26] can also be obtained by modifying the existing enzyme using mutagenesis via directed evolution, rational or semi-rational approaches [27–35]. Random mutagenesis of a single gene can be done by chemical, error prone-PCR or saturation mutagenesis, or by using mutator strains. On the other hand, gene recombination approach can be applied with more than one related gene sequences, using tools like DNA shuffling, Random Chimeragenesis on Transient Templates (RACHITT), Exon shuffling, incremental truncation for the creation of hybrid enzymes (ITCHY), Sequence Homology-Independent Protein Recombination (SHIPREC). The reader is referred to review by Rubin-pital et al. [31] for details about these processes, their advantages and drawbacks. Recent developments along with additions of rational component have resulted in faster selection methods and maximized qualities of libraries with more relevant mutations [36]. Rational mutagenesis to improve enzyme property has been attempted in recent years to obtain the desired property; however, the phenotype of certain mutations is still beyond the current understanding of enzyme structure and function.

Conclusion

In the past few years, enzyme production and research have taken a major leap and a vast number of potential enzymes are available in market and are produced at industrial scale. Reports are being continuously published related to the screening and finding newer and better enzymes. However, it is generally observed that wild-type enzymes are not directly applicable for an industrial process. In the coming years, it is expected that more industrially important enzymes will be discovered or engineered that can satisfy the ever-growing demand of enzymes. The availability of various expression vectors, host and systems has increased the possibility of expressing a gene artificially in a host of our choice. However, protein expression even in bacterial host like *E. coli* can be challenging many times [37, 38]. The diversity of enzymes present in databases indicates that the present knowledge of structure and function is vast but far from complete. The last two decades has seen

tremendous growth in protein structural information, and expression systems and tools have enriched in large, but we still require more information to understand and utilize it to its full potential. With the rise in molecular techniques, enzyme improvement by protein engineering has taken a big leap [35]. Drastic improvement in enzymatic properties like activity and stability has been witnessed by using methods of directed evolution or rational mutagenesis. With the current knowledge of enzyme structure and function, it is still a challenging task to pursue a rational approach of enzyme engineering in every case to improve their properties. Efforts should be more focussed towards solving enzyme crystal structures and expanding our knowledge and understanding of enzyme function and properties. The pace of structure information cannot be compared with the way new genes or proteins are being discovered, but attempts can be made to improve it further. Generating and analysing diverse crystallographic data will help in understanding the enzymes in greater details, and also, will help in rational engineering of the enzyme for improved properties. There is an urgent demand for developing new tools and pipelines which can handle and analyse the exponentially growing database, and related experimental literature, with minimal manual intervention. This will help in discovering novel and better enzymes comparatively faster with high success rate.

Compliance with Ethical Standards

Conflict of interest The authors declare no conflict of interest.

Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

Informed Consent For this type of study, formal consent is not required.

References

1. Coker, J. A. (2016). Extremophiles and biotechnology: Current uses and prospects. *F1000Research, 5*, 396.
2. Anastas, P. T., & Warner, J. C. (1998). *Green chemistry: Theory and practice*. Oxford University Press: New York.
3. Savile, C. K., Janey, J. M., Mundorff, E. C., Moore, J. C., Tam, S., Jarvis, W. R., … Hughes, G. J. (2010). Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. *Science, 329*(5989), 305–309.
4. Chamoli, M., & Pant, K. (2014). In-silico bioprospecting of the enzymes involved in the biosynthetic pathway of the alkaloid berberine and its distance study Through R. *International Journal of Advanced Technology in Engineering and Science, 2*(9), 165–178.
5. Musumeci, M. A., Lozada, M., Rial, D. V., Cormack, W. P. M., Jansson, J. K., Sjöling, S., … Dionisi, H. M. (2017). Prospecting biotechnologically-relevant monooxygenases from cold sediment metagenomes: An in silico approach. *Marine Drugs, 15*(4).
6. Tan, H., Wu, X., Xie, L., Huang, Z., Peng, W., & Gan, B. (2016). Identification and characterization of a mesophilic phytase highly resilient to high-temperatures from a fungus-garden associated metagenome. *Applied Microbiology and Biotechnology, 100*(5), 2225–2241.
7. Berón, C. M., Curatti, L., & Salerno, G. L. (2005). New strategy for identification of novel cry-type genes from bacillus thuringiensis strains. *Applied and Environmental Microbiology, 71*(2), 761–765.
8. Tan, H., Wu, X., Xie, L., Huang, Z., Peng, W., & Gan, B. (2016). A novel phytase derived from an acidic peat-soil microbiome showing high stability under acidic plus pepsin conditions. *Journal of Molecular Microbiology and Biotechnology, 26*(4), 291–301.
9. Shakeel, T., Gupta, M., Fatma, Z., Kumar, R., Kumar, R., Singh, R., … Yazdani, S. S. (2018). A consensus-guided approach yields a heat-stable alkane-producing enzyme and identifies residues promoting thermostability. *The Journal of Biological Chemistry*, 1–30.
10. Sharma, N., Thakur, N., Raj, T., Savitri, & Bhalla, T. C. (2017). Mining of microbial genomes for the novel sources of nitrilases. *BioMed Research International, 2017*.
11. Gupta, S., Singh, Y., Kumar, H., Raj, U., Rao, A. R., & Varadwaj, P. K. (2018). Identification of novel abiotic stress proteins in triticum aestivum through functional annotation of hypothetical proteins. *Interdisciplinary Sciences: Computational Life Sciences, 10*(1), 205–220.
12. Thornbury, M., Sicheri, J., Guinard, C., Mahoney, D., Routledge, F., Curry, M., … Getz, L. (2018). Discovery and Characterization of Novel Lignocellulose-Degrading Enzymes from the Porcupine Microbiome. *bioRxiv*, (February).
13. Toyama, D., de Morais, M. A. B., Ramos, F. C., Zanphorlin, L. M., Tonoli, C. C. C., Balula, A. F., *et al*. (2018). A novel β-glucosidase isolated from the microbial metagenome of Lake Poraquê (Amazon, Brazil). *Biochimica et Biophysica Acta, 1866*(4), 569–579.
14. Qu, Y., Egelund, J., Gilson, P. R., Houghton, F., Gleeson, P. A., Schultz, C. J., & Bacic, A. (2008). Identification of a novel group of putative Arabidopsis thaliana β-(1,3)-galactosyltransferases. *Plant Molecular Biology, 68*(1–2), 43–59.
15. Foong, C. P., Lakshmanan, M., Abe, H., Taylor, T. D., Foong, S. Y., & Sudesh, K. (2018). A novel and wide substrate specific polyhydroxyalkanoate (PHA) synthase from unculturable bacteria found in mangrove soil. *Journal of Polymer Research, 25*(1), 23.
16. Vaquero, M. E., De Eugenio, L. I., Martínez, M. J., & Barriuso, J. (2015). A novel CalB-type lipase discovered by fungal genomes mining. *PLoS ONE, 10*(4), 1–11.
17. Adam, N., & Perner, M. (2018). Novel hydrogenases from deep-sea hydrothermal vent metagenomes identified by a recently developed activity-based screen. *ISME Journal, 12*(5), 1225–1236.
18. Ferrer, M., Martínez-Martínez, M., Bargiela, R., Streit, W. R., Golyshina, O. V., & Golyshin, P. N. (2016). Estimating the success of enzyme bioprospecting through metagenomics: Current status and future trends. *Microbial Biotechnology, 9*(1), 22–34.
19. Uria, A. R., & Zilda, D. S. (2016). Metagenomics-guided mining of commercially useful biocatalysts from marine microorganisms. In Advances in Food and nutrition research.
20. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server. *The Proteomics Protocols Handbook*, 571–607.
21. Roumpeka, D. D., Wallace, R. J., Escalettes, F., Fotheringham, I., & Watson, M. (2017). A review of bioinformatics tools for bio-prospecting from metagenomic sequence data. *Frontiers in Genetics*.

22. Machielsen, R., Leferink, N. G. H., Hendriks, A., Brouns, S. J. J., Hennemann, H. G., Daußmann, T., & Van Der Oost, J. (2008). Laboratory evolution of Pyrococcus furiosus alcohol dehydrogenase to improve the production of (2S,5S)-hexanediol at moderate temperatures. *Extremophiles, 12*(4), 587–594.

23. Wang, N. Q., Sun, J., Huang, J., & Wang, P. (2014). Cloning, expression, and directed evolution of carbonyl reductase from Leifsonia xyli HS0904 with enhanced catalytic efficiency. *Applied Microbiology and Biotechnology, 98*(20), 8591–8601.

24. Jakoblinnert, A., Wachtmeister, J., Schukur, L., Shivange, A. V., Bocola, M., Ansorge-Schumacher, M. B., & Schwaneberg, U. (2013). Reengineered carbonyl reductase for reducing methyl-substituted cyclohexanones. *Protein Engineering, Design and Selection, 26*(4), 291–298.

25. Hoelsch, K., Sührer, I., Heusel, M., & Weuster-Botz, D. (2013). Engineering of formate dehydrogenase: Synergistic effect of mutations affecting cofactor specificity and chemical stability. *Applied Microbiology and Biotechnology, 97*(6), 2473–2481.

26. Koudelakova, T., Chaloupkova, R., Brezovsky, J., Prokop, Z., Sebestova, E., Hesseler, M., … Damborsky, J. (2013). Engineering enzyme stability and resistance to an organic cosolvent by modification of residues in the access tunnel. *Angewandte Chemie - International Edition, 52*(7), 1959–1963.

27. Buller, A. R., Brinkmann-Chen, S., Romney, D. K., Herger, M., Murciano-Calles, J., & Arnold, F. H. (2015). Directed evolution of the tryptophan synthase β-subunit for stand-alone function recapitulates allosteric activation. *Proceedings of the National Academy of Sciences, 112*(47), 14599–14604.

28. Brinkmann-Chen, S., Flock, T., Cahn, J. K. B., Snow, C. D., Brustad, E. M., McIntosh, J. A., … Arnold, F. H. (2013). General approach to reversing ketol-acid reductoisomerase cofactor dependence from NADPH to NADH. *Proceedings of the National Academy of Sciences, 110*(27), 10946–10951.

29. Fox, R. J., & Huisman, G. W. (2008). Enzyme optimization: moving from blind evolution to statistical exploration of sequence-function space. *Trends in Biotechnology, 26*(3), 132–138.

30. Reetz, M. T., Rentzsch, M., Pletsch, A., Maywald, M., Maiwald, P., Peyralans, J. J. P., … Taglieber, A. (2007). Directed evolution of enantioselective hybrid catalysts: a novel concept in asymmetric catalysis. *Tetrahedron, 63*(28), 6404–6414.

31. Rubin-Pitel, S. B., Cho, C. M. H., Chen, W., & Zhao, H. (2007). Directed evolution tools in bioproduct and bioprocess development. *Bioprocessing for Value-Added Products from Renewable Resources*, 49–72.

32. Li, Y., & Cirino, P. C. (2014). Recent advances in engineering proteins for biocatalysis. *Biotechnology and Bioengineering, 111*(7), 1273–1287.

33. Wang, M., Si, T., & Zhao, H. (2012). Biocatalyst development by directed evolution, *Bioresource Technology, 40*(6), 1301–1315.

34. Woodley, J. M. (2013). Protein engineering of enzymes for process applications. *Current Opinion in Chemical Biology, 17*(2), 310–316.

35. Zheng, G. W., & Xu, J. H. (2011). New opportunities for biocatalysis: Driving the synthesis of chiral chemicals. *Current Opinion in Biotechnology, 22*(6), 784–792.

36. Lane, M. D., & Seelig, B. (2014). Advances in the directed evolution of proteins. *Current Opinion in Chemical Biology, 22*, 129–136.

37. Kaur, J., Kumar, A., & Kaur, J. (2018). Strategies for optimization of heterologous protein expression in *E. coli*: Roadblocks and reinforcements. *International Journal of Biological Macromolecules, 106*, 803–822.

38. Rosano, G. L., & Ceccarelli, E. A. (2014). Recombinant protein expression in *Escherichia coli*: Advances and challenges. *Frontiers in Microbiology*.