Life Science Informatics Publications

Research Journal of Life Sciences, Bioinformatics, Pharmaceutical and Chemical Sciences

Journal Home page http://www.rjlbpcs.com/

RJLBPCS

ISSN 2454-6348

Original Review Article **DOI - 10.26479/2017.0205.10**

COMPUTATIONAL METHODS: TOOLS FOR STRUCTURE-FUNCTIONAL STUDIES AND PROTEIN ENGINEERING

Suseela Lanka [1]*, Muralidhar Pydipalli [2]

1. Department of Biotechnology, Krishna University, Machilipatnam.

2. VIMTA Labs, Cherlapally, Hyderabad

ABSTRACT: In view of an increased demand for microbial originated Industrial enzymes in a wide variety of fields such as food, dairy, pharmaceutical, detergent, textile, oleo-chemical, perfume and cosmetic industries etc., there is every need to improve the stability, specificity and catalytic efficiencies of the enzymes to obtain better yields which can be possible through protein engineering using various computational methods. Availability of genomic and proteomic sequence databases resulted in an explosion of information which aid in discovering new microbial enzymes whose properties can be improved or modified. The present review mainly focuses on various computational tools for structure prediction (homology modeling) and protein engineering.

KEYWORDS: Industrial enzymes, catalytic efficiencies, computational methods, structure prediction, protein engineering.

***Corresponding Author: Dr. Suseela Lanka** Ph.D.

Department of Biotechnology, Krishna University Machilipatnam – 521001, Andhra Pradesh, India

* Email Address: susheelalankaku@gmail.com

1. INTRODUCTION

Enzymes, the biological catalysts are considered as nature's gifts. Microbial enzymes are widely used as biological catalysts because of their immense use in various industrial applications [1]. More than 500 different industrial products are being made using various microbial enzymes [2, 3]. Chemical catalyzed reactions have several limitations like – low catalytic efficiency, often the reactions need to

be carried out at higher temperatures and pressures and even they don't exhibit enantio- specificity for synthesis of chiral drugs. In addition these reactions need to be carried out in the presence of organic solvents which ultimately results in the formation of organic waste and there by contributes to pollution. The enzyme catalyzed reactions in other way overcomes many of these limitations as biological enzymes works under milder conditions and have high stereo selectivity and broader substrate specificities [4]. But the only drawback with the enzymes is that they only work under the conditions that are confined to them. A large scale industrial application requires the use of enzymes with improved catalytic efficiency, greater stability and specificity. This problem can be rectified by employing various computational methods that engineer the protein to modify them to suit for industrial needs. The present review mainly focuses on various available computational tools for 3D modeling and protein engineering.

GENOMIC AND PROTEOMIC DATABASES

Advances in genome sequencing programs resulted in an explosion of information made available from sequence databases and searching of these databases will explore novel enzymes. This can be done by using data mining [5]. Till now there are more than 2000 genome sequences available in the NCBI database (http://www.ncbi.nlm.nhi.gov/genomes). Bioinformatics also provides various searching tools (Ex. BLAST), which provides homologous sequences based on search for conserved regions by aligning the sequences. These sequences can be used for further characterization.

BIOINFORMATICS APPROACHES FOR STRUCTURE ANALYSIS

The primary prerequisite for protein analysis is to understand the structure of the protein. Structural information further provides insight into the mechanism of action of the enzymes. Sequence analysis and structure prediction of proteins can be done using bioinformatics tools. The sequence information regarding proteins will be stored in protein databases. Primary databases stores information about protein sequences whereas the secondary databases contains the information pertaining to the analysis of the sequences of primary databases. Table 1 shows the various primary and secondary databases. SCOP (Structural Classification Of Proteins) [6], CATH (Class, Architecture, Topology and Homology) and PDB Sum [7] are the major protein structure classification databases. Table 2 provides information about various available tools for protein structure determination. Primary structure analysis of the protein can be done by using Prot Param Protein Engineering and Bioinformatics tool of EXPASY [8] which provides information about the protein sequence (no of amino acids), Molecular weight of the protein, pI, composition of amino acids, grand hydrophobicity etc. Secondary structure analysis can be done using SOPMA; PROSITE tool provides information about protein family, active site of the protein etc. There are various tools available for homology

modelling (Ex: SWISS MODEL) [9, 10], ProMod [11] to obtain 3D structure of the protein. Once the model is obtained there are also tools to check for the quality of the model like- PROCHECK [12], ProSA [13] etc.

PROTEIN ENGINEERING AND BIOINFORMATICS

Enzymes as natural catalysts are widely involved in various applications. In olden days, the applications of the enzymes are restricted or confined around to the limitations of the enzymes but now a day with the advances in the protein engineering one can engineer the enzymes to fit for the required process [14]. Protein Engineering mainly aims at construction of enzymes with improved activities, enzymes with greater stability and enhanced specificities. Table 3 shows the various computational tools that were used to improve enzyme activity, stability and specificity.

DISULFIDE ENGINEERING IN PROTEINS TO IMPROVE THERMAL STABILITY

It was a known fact that disulfide bonds provide stability to many extracellular and secreted proteins. Disulfide bonds are believed to increase the stability of the protein by decreasing the conformational entropy and by raising the free energy of the denatured state [15]. The protein stability and function can be drastically improved by introducing novel disulfide bonds into the protein. Disulfide by Design 2.0 (DbD2) is a computational application which facilitates the identification of potential disulfides that are not only likely to form but are also expected to provide improved thermal stability to the protein. A web server hosting DbD2 is provided at http://cptweb.cpt.wayne.edu/DbD2/

2. CONCLUSION

The present review provides information about primary and secondary databases for protein analysis, tools for 3D modeling of proteins & various computational tools and their applications in Protein Engineering etc. The knowledge of computational tools helps one to engineer the proteins to modify them to suit for industrial needs.

List of Tables

Table 1: Primary and Secondary databases for protein analysis

Type of databases	Databases	Targets	Web address
Primary	PIR	Sequence	http://pir.georgetown.edu
	MIPS	Sequence	http://mips.gsf.de
	Swiss-Prot	Sequence	http://www.ebi.ac.uk/swissprot/
Secondary	PROSITE	Patterns	http://www.expasy.ch/prosite/
	PRINTS	Finger prints	http://www.bioinf.manchester.
	Pfam	HMM, MMS	http://pfam.sanger.ac.uk/
	BLOCKS	Motifs	http://blocks.fhcrc.org/

Table 2: Tools for 3D Modeling

Software tool	Method
HH Pred	Template detection, alignment, 3D modeling
Raptor X	Remote homology detection, Protein 3D modeling, binding site prediction
MODELLER	Satisfaction of spatial restraints
SWISS-MODEL	Local similarity/fragment assembly
Phyre and Phyre 2	Remote template detection, alignment, 3D modeling, multi templates, *ab initio* template detection
Esy Pred 3D	Templates, *ab initio* template detection, alignment, 3D modeling
ROBETTA	Rosetta homology modeling, *ab initio* fragment assembly with Ginzu domain prediction
BHAGEERATH	Combination of *ab initio* folding and homology method

Table 3: Computational tools and their applications in Protein Engineering

Computational tools	Utility	References
ZEBRA	Mainly identifies and analyses the conserved SSP'S (Subfamily Specific Positions) by analyzing enzyme functional subfamilies.	[16]
JANUS	Predicts mutations by analyzing multiple sequence alignments which are required for interconversion of structurally related but functionally distinct enzymes.	[17]
CAVER	Used to analyze tunnel dynamics by molecular dynamics simulations.	[18]
ROSETTA and ORBIT	Web based applications widely sued for denovo designing of the proteins.	[19, 20, 21]
SABER	Analyses the functional sites of the proteins stored in the PDB.	[22]
ASRA	Employed for directed evolution experiments. It identifies the underlying regularity of the protein property landscape and makes predictions about the properties of uncharacterized proteins	[23]
Empirical Valence bond (Molecular modeling)	Modeling technique used for a quantitative analysis of enantioselectivity of enzymes.	[24]
PRETHERMUT	To evaluate the free energy of unfolding.	[25]
POCKETOPTIMIZER	used to modify the residues making up the protein binding pocket to improve or newly establish the binding of a small ligand by making use of the molecular modeling programs AUTO DOCK VINA and CADDSUITE, and AMBER to analyse protein packing.	[26]
CCSM (combinatorial coevolving-site saturation mutagenesis)	Method for identifying hotspots for mutagenesis.	[27]

ROSETTA VIP	It uses the ROSETTA HOLES analysis module and a simple geometric scoring function to identify a small set of mutations that may yield improved packing. The protocol is applicable for stabilizing both designed and native proteins against chemicals and thermal denaturation.	[28]
WISDOM	tool enables searching for templates, designing optimized sequences with stability, analyzing fold specificity and binding affinity, and quantitative assessment of the designs by ranking of sequences as well as structures	[29]
EVODESIGN	A web-based tool for designing optimal protein sequences of given scaffolds while predicting multiple sequence and structure-based features for design ranking.	[30]
PYROSETTA	Graphical user interface for preparing and running protocols of ROSETTA and for data analysis.	[31]

REFERENCES

1. Adrio JL, Demain AL. Microbial Cells and Enzymes—A Century of Progress. In Methods in Biotechnology. Microbial Enzymes and Biotransformations. Barredo JL. Ed.; Vol. 17, Humana Press, Totowa, NJ, USA, 2005, pp. 1–27.
2. Johannes TW. Zhao, H. Directed evolution of enzymes and biosynthetic pathways. Curr Opin Microbiol. 2006; 9: 261–267.
3. Kumar A, Singh S. Directed evolution: Tailoring biocatalysis for industrial application. Crit Rev Biotechnol. 2013; 33: 365–378.
4. Johnson EA. Biotechnology of non-*Saccharomyces* yeasts- the ascomycetes. Appl Microbiol Biotechnol.2013; 97: 503–517.
5. Van Lanen S, Shen B. Microbial genomics for the improvement of natural product discovery. Curr Opin Microbiol. 2006; 9: 252–260.
6. Murzin AG, Brenner SE, Hubbard T et al. Scop: a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol. 1995; 247: 536–40.
7. Laskowski RA, Hutchinson EG, Michie AD et al. PDBSum: a web based database of summaries and analysis of all PDB structures. TIBS. 1997; 22: 488–90.
8. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. ExPASy: The proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res. 2003; 31 (13): 3784–8.
9. Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL Workspace: A web-based environment for protein structure homology modeling. .Bioinformatics. 2006; 22: 195-201.
10. Marco Biasini, Stefan Bienert, Andrew Waterhouse, Konstantin Arnold, Gabriel Studer, Tobias Schmidt, Florian Kiefer, Tiziano Gallo Cassarino, Martino Bertoni, Lorenza Bordoli, Torsten Schwede. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research*; 2014; 42 (W1): W252-W258.
11. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modelling. Electrophoresis. 1997; 18: 2714-2723.

12. Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR. 1996; 8(4): 477-86.
13. Widerstein, Sippl. ProSA-web: interactive web service for the recognition of errors in three-dimensional structure of proteins. Nucleic Acids Res. 2001; 35: W401-W410.
14. Bornscheuer UT, Huisman GW, Kazlauskas RJ, Lutz S, Moore JC, Robins K. Engineering the third wave of biocatalysis. Nature. 2012; 485: 185-194.
15. Flory PJ. Theory of elastic mechanisms in fibrous proteins. J Am Chem Soc. 1956; 78(20): 5222–5235.
16. Suplatov D, Kirilin E, Takhaveev V, Svedas V. Zebra: a web server for bioinformatic analysis of diverse protein families. J Biomol Struct Dyn. 2013; 32(11): 1752-1758.
17. Addington TA, Mertz RW, Siegel JB, Thompson JM, Fisher AJ, Filkov V, Fleischman NM, Suen AA, Zhang C, Toney MD. Janus: prediction and ranking of mutations required for functional interconversion of enzymes. J Mol Biol. 2013; 425: 1378-1389.
18. Chovancova E, Pavelka A, Benes P, Strnad O, Brezovsky J, Kozlikova B, Gora A, Sustr V, Klvana M, Medek P et al. CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures. PLoS Comput Biol. 2012; 8:e1002708.
19. Hilvert D. Design of protein catalysts. Annu Rev Biochem. 2013; 82: 447-470.
20. Khare SD, Fleishman SJ. Emerging themes in the computational design of novel enzymes and protein–protein interfaces. FEBS Lett. 2013; 587: 1147-1154.
21. Kries H, Blomberg R, Hilvert D. De novo enzymes by computational design. Curr Opin Chem Biol. 2013; 17: 221-228.
22. Nosrati GR, Houk KN. SABER: a computational method for identifying active sites for new reactions. Protein Sci. 2012; 21: 697-706.
23. Feng X, Sanchis J, Reetz MT, Rabitz H. Enhancing the efficiency of directed evolution in focused enzyme libraries by the adaptive substituent reordering algorithm. Chem Eur J. 2012; 18: 5646-5654.

Peer review under responsibility of Life Science Informatics Publications
2017 Jan- Feb RJLBPCS 2(5) Page No.117

24. Frushicheva MP, Warshel A. Towards quantitative computer-aided studies of enzymatic enantioselectivity: the case of Candida antarctica lipase A. Chem BioChem. 2012; 13: 215-223.
25. Tian J, Wang P, Huang L, Chu X, Wu N, Fan Y. Improving the thermostability of methyl parathion hydrolase from Ochrobactrum sp. M231 using a computationally aided method. Appl Microbiol Biotechnol. 2013; 97: 2997-3006.
26. Malisi C, Schumann M, Toussaint NC, Kageyama J, Kohlbacher O, Ho¨ cker B. Binding pocket optimization by computational protein design. PLoS ONE. 2012; 7:e52505.
27. Wang C, Huang R, He B, Du Q. Improving the thermostability of alpha-amylase by combinatorial coevolving-site saturation mutagenesis. BMC Bioinformatics. 2012; 13: 263.
28. Borgo B, Havranek JJ. Automated selection of stabilizing mutations in designed and natural proteins. Proc Natl Acad Sci. USA. 2012; 109: 1494-1499.
29. Smadbeck J, Peterson MB, Khoury GA, Taylor MS, Floudas CA. Protein WISDOM: a workbench for in silico de novo design of biomolecules. J Vis Exp. 2013; 77: e50476. http://dx.doi.org/10.3791/50476.
30. Mitra P, Shultis D, Zhang Y. EvoDesign: de novo protein design based on structural and evolutionary profiles. Nucleic Acids Res. 2013; 41: W273-W280.
31. Adolf-Bryfogle J, Dunbrack RL Jr. The PyRosetta Toolkit: a graphical user interface for the Rosetta software suite. PLoS ONE. 2013; 8:e66856.
32. NCBI Microbial Genomes. http://www.ncbi.nlm.nhi.gov/genomes/microbial (Accessed on 10 July 2013)