iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 **Issue:** V **Month of publication:** May 2018

DOI: http://doi.org/10.22214/ijraset.2018.5312

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue V, May 2018- Available at www.ijraset.com

In Silico Identification and Characterization of Potent Drug Targets in Fungal Pathogen causing Keratitis

Kumari Rashmi[1], Himanshu A. Pandya[2] and Rakesh M. Rawal[3]
[1]*Department of Botany, Bioinformatics and Climate Change Impact Management, Ahmedabad -380009, Gujarat, India.*
[2]*Department of Botany, Bioinformatics and Climate Change Impact Management, Ahmedabad -380009, Gujarat, India.*
[3]*Department of life Sciences, Ahmedabad -380009, Gujarat, India*

Abstract: ***Availability of genome sequences of pathogens has provided a tremendous amount of information that can be useful in drug target identification. Subtractive proteomics approach is being used to mine the list of proteins present in different Aspergillus species which are non-homologous to human and essential for the survival of the pathogen. Current study is based on complete proteome information of Aspergillus terreus strain NIH 2624available in biological databases to identify putative protein targets. Subjecting the total set of pathogen non homologous proteins (10402) against the Database of Essential Genes 2030 proteins were screened out as essential proteins of the Aspergillus terreus,out of which 136 are found to be essential in the pathogen with no human homolog. Among 136 proteins; around 88 proteins were found to be putative uncharacterized using CELLO. The hypothetical essential proteins were functionally annotated through SVMProt server. Druggability of each of the identified drug targets was also evaluated by the Drug Bank database.***
Keywords: Aspergillus terreus, BLAST, DEG, Drug targets, Subtractive proteomics

I. INTRODUCTION

Fungal Keratitis is a fungal infection of the cornea. It primarily affects the corneal epithelium and stroma, although the endothelium and anterior chamber of the eye may get involved in more severe disease [1]. Aspergillus terreus is a ubiquitous fungus in our environment [2]. Aspergillus terreus is an economically important species from a number of aspects [3-4]. In silico drug target discovery is now possible primarily because of the Human Genome Project and related large-scale sequencing efforts. These data are allowing researchers to identify literally thousands of drug targets for both endogenous diseases and infectious disease. Many of these targets are being, or can be rapidly identified in silico using simple sequence comparison and sequence alignment software [5]. Selecting new molecular targets by comparative genomics, homology modeling and virtual screening of compounds is promising in the process of new drug discovery [6]. Subtractive proteomic approach helps in the identification of essential genes of pathogen which must not be homologous to the host genome, because homologous protein target may create problems of self-binding. These non-homologous essential genes ensure the survival of the pathogen and therefore can be targeted for drug development [7]. This subtractive genomics approach has been successfully used to identify novel drug targets in several pathogens such as Pseudomonas aeruginosa [8], Helicobacter pylori [9], Brugiamalayi [10], Campylobacter fetus [11], Leptospira interrogans [12] and Mycobacterium ulceran [13].

The current study on Aspergillus terreus is based on proteome subtraction approach. Differential pathway analysis, sub cellular localization prediction and functional family classification of putative uncharacterized essential proteins were done to analyze the complete proteome of Aspergillus terreus to find out potential vaccine candidates and therapeutic drug targets. Subtractive proteomics analysis helps to subtract essential proteins from non-target proteins in the pathogen and ensures the treatment of disease with fewer side effects to the host. The application of the database of essential genes helps to identify the potential drug targets in pathogens. Protein sequences from different strains of a single organism offer comprehensive information for determining the genetic basis of biological functions and illustrating the metabolic mechanisms. This information can be significantly used to identify global protein targets in the pathogen.

II. MATERIAL AND METHODS

In the present investigation, multidrug resistant Aspergillus terreus strain NIH 2624 was selected. The proteome of Aspergillus terreus strain NIH 2624 was downloaded from UniProt (www.uniprot.org) in fasta format. The proteins having sequence lengths

less than 100 amino acids were eliminated as they are less likely to represent essential genes. This was on the assumption that proteins less than 100 amino acids in length were unlikely to represent essential proteins, yet be unique to the organism.

A. Removal of Duplicate Protein

Pathogens proteome were purged at 60% threshold using CD-HIT to exclude paralogs or duplicates within the proteome of *Aspergillus terreus* strain NIH 2624 [14]. The paralogs were excluded and the remaining sets of protein were used for further analysis.

B. Identification of Essential Proteins

The selected clustered proteins were then subjected to similarity search using BlastP against Database of Essential Genes (DEG) with E-value 10^{-4} and Bits score greater than 100 to screen out essential gene proteins [15] [16] [17].

C. Sequence Homology with Human Proteome

The set of essential proteins obtain were subjected to BlastP against Homo sapiens proteome with the expectation value (E-value) cut-off of 10^{-4}. The human homologous were excluded and the lists of non-homologs were compiled.

D. Sub-cellular Localization Prediction

Sub-cellular localization analysis of the essential proteins has been done by cello to identify the surface membrane proteins which could be probable vaccine targets [18].

E. Functional Family Prediction of Putative Uncharacterized Protein

Functional family prediction of the putative uncharacterized essential proteins was done by using the SVMProt web server [19].

F. Metabolic Pathway Analysis

Metabolic pathway analysis of the essential proteins was done by KAAS server at KEGG for the identification of potential targets. KAAS (KEGG Automatic Annotation Server) provides functional annotation of genes by BLAST comparisons against the manually curated KEGG GENES database. Comparative analysis of the metabolic pathways of the host and pathogen was performed by using Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database to trace out essential proteins involved in pathogen specific metabolic pathways for the identification of potential drug targets [20-21].

G. Druggability Analysis

Druggability analysis was performed for each of the short-listed targets using various databases. Each of the targets was subjected to a homology search against Drug Bank, TTD, PDTD and HIT databases [22] [23] [24] and [25].

H. Interactome Analysis

A protein-protein interaction network was constructed for each of the short-listed novel targets using STRING 9.0 a precomputed database for the exploration of protein–protein interactions. Using string high confidence interactors for each novel target protein in *Aspergillus terreus* were predicted [26].

III. RESULT AND DISCUSSION

The complete proteome sequence of Aspergillus terreus NIH 2624 strain retrieved from uniprot comprised of 10402 protein sequences. Results of subtractive proteomics approach are summarized in the Table 1.

Total 10308 proteins were found after filtering of sequences. Then sequences were purged at 60% using CD-HIT to remove paralogues, respectively. Out of 10308 proteins were found duplicate or paralogs proteins. Thereafter 205 paralogs were excluded and resulting 10103 protein sequences were run through the database of essential genes (DEG) at an expectation cut-off of 10^{-4}, yielding 2030 essential genes respectively.

Those 2030 essential proteins identified were subjected to BLASTP against the human genome proteins to exclude any proteins that have a significant match (E-value cutoff of 10^{-4}) with human homologs; the essential, non-human homologs were listed out. Consensually, 136 proteins respectively were considered as having no close relatives in human. Thus, these 136 proteins might be considered as potential and unique drug targets. Among 136 protein sequences 88 were found to be putative and uncharacterized proteins.

Table 1 Summary of subtractive proteomic analysis of *Aspergillus terreus NIH 2624*

Summary of Analysis	Result
Total Number of Proteins	10402
Total protein after elimination of Proteins having <100 amino acids	10308
After removal of Duplicates proteins (>60% identical) in CD-HIT	10103
Essential protein in DEG (E-value 10^{-4})	2030
Non-human homologous proteins (E-value 10^{-4})	136
Putative uncharacterized proteins	88

Based on the localization score, CELLO sorted the location of 17 as nuclear, 33 as cytoplasmic, 52 as plasmamembrane, 18 as mitochondrial, 10 as chloroplast, 4 as extracellular and 2 as peroxisomal (Fig. 1).

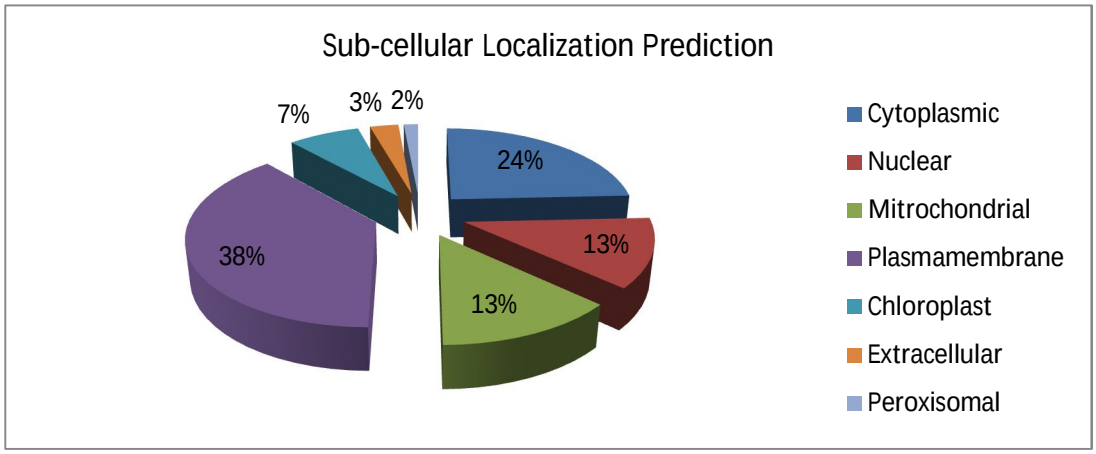


Fig. 1 Percentage distribution of sub-cellular locations in *Aspergillus terreus*

The functional classification of these 88 uncharacterized essential proteins was performed using SVM-Prot web server (Fig. 2). The web server classifies 88 proteins as transmembrane proteins (27), lipid-binding proteins (13), zinc-binding proteins(13), iron-binding proteins (12), electrochemical potential driven transporter proteins (10), DNA- binding proteins (1), repressorproteins (1), metal-binding proteins(1), transferases (2), hydrolases (1),G protein coupled receptor (1), repressor (1), lyases (1), oxidoreductase (1), structural proteins(1) and two proteins classified with very low p value (58.6%) among which one protein involved in photosystem 1 other in calcium, magnesium-binding.

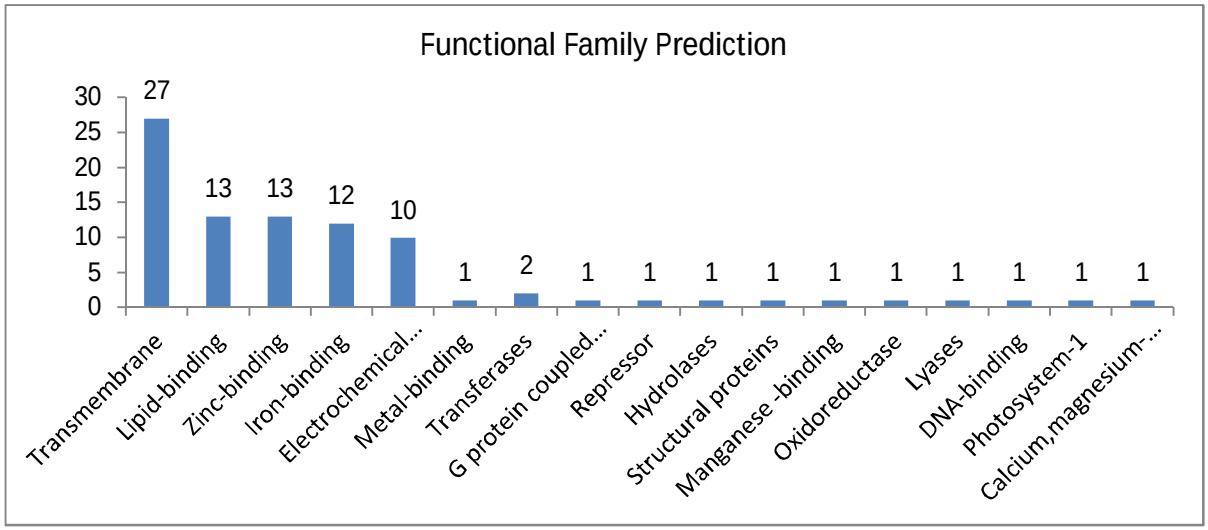


Fig. 2 Functional family prediction of uncharacterized proteins

Metabolic pathway analysis of these 136 non-human homologous essential proteins revealed that 91 proteins are involved in metabolic pathway, among these 4 pathways were unique to pathogen these unique pathways were Methane metabolism, Styrene metabolism, Carbon fixation and photosynthetic organisms and two component system.

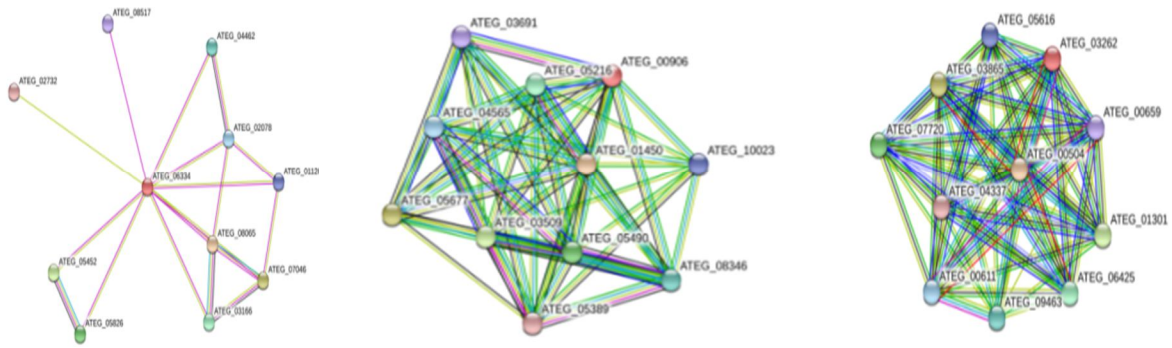
A 'druggable' target should have potential to bind to the drug-like molecules with high affinity. In the current approach, the druggability of the short-listed potential targets was evaluated by sequence similarity search against targets from DrugBank, HIT, PDTD and TTD databases. By a similar homology search against these databases aligned targets were considered as significant homologs. All the other targets showing no similarity with these drug target databases were differentiated as novel targets, which should be further validated experimentally [27]. In order to recognize the important role of chemical accessibility in ranking of putative drug targets we looked for druggability information on the predicted essential targets. Druggability describes the properties of a protein that make them able to interact with a drug-like molecule [28]. In *Aspergillus terreus* NIH 2624 nine novel proteins (Table 2) were found for further laboratory as well as commercial use. Thus, in order to establish the reliability of data further in vivo or in vitro assessment of the predicted proteins is needed.

Table 2 Novel proteins after druggability analysis

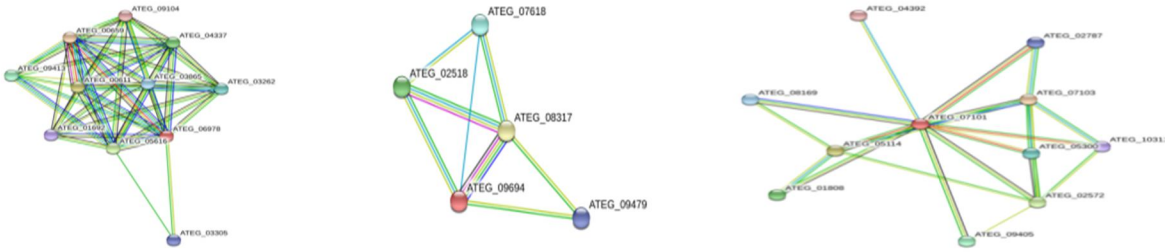
S. No.	Novel proteins in *Aspergillus Terreus* NIH 2624
1	ROT1_ASPTN Protein rot1
2	Homocitrate synthase
3	Anthranilate synthase component I
4	Imidazoleglycerol-phosphate dehydratase
5	Glutamine amidotransferase subunit pdxT
6	Biotin synthase
7	Dihydroxy-acid dehydratase
8	Cyanide hydratase
9	Thiamine-phosphate pyrophosphorylase

Protein network analysis is generally used for selection of drug targets having interaction with maximum numbers of other targets, an efficient technique to discover a potent drug target against pathogens [19]. In STRING, one protein is interacted with a number of proteins and showed the strength of interaction as score. The interacting score depends on Neighborhood in the genome, Gene fusions, Co-occurrence across genomes, Co-Expression, Association in curated databases and text mining [29]. High confidence interactors with score greater than or equal to 0.700 alone were included in the protein network and predicted the functional interactors of the query target (Figure. 3) (Table 3). The analysis of STRING data showed that all the proposed drug targets showed interactions with the high confidence score. Based on the variation in the critical network parameter values the potentiality of the targets was determined. Target protein interacting with more proteins is considered as metabolically important active protein, which can act as an appropriate drug target [30-31].

(a) Q0CJ00 (b) Q0CZH8 (c) Q0CSS2



(d)Q0CH64 (e)Q0C9E0 (f)Q0CGU1



(g)Q0CZ04 (h) Q0CSX2 (i)Q0C8V5

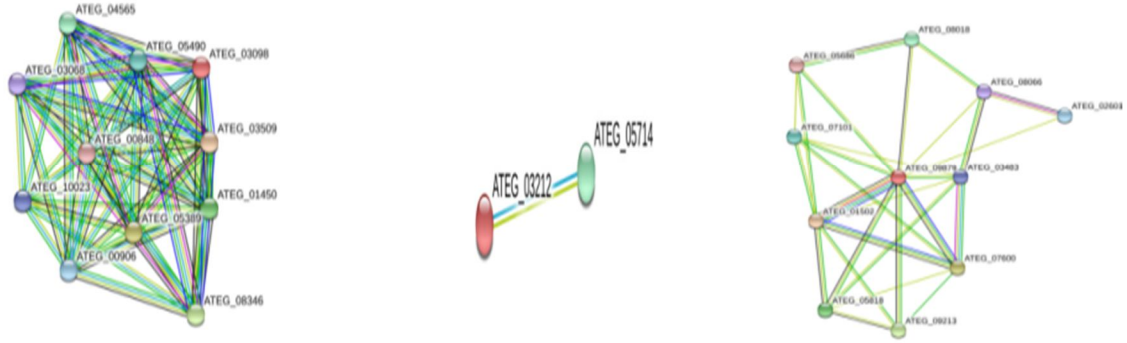


Fig. 3 The protein-protein network study of drug targets namely (a)Protein rot 1 (b)Homocitrate synthase (c)Anthranilate synthase component I (d)Imidazole glycerol-phosphate (e)Glutamine amidotransferase subunit pdXT (f)Biotin synthase dehydratase (g)Dihydroxy-acid dehydratase (h)Cynide hydratase (i)Thiamine-phosphate pyrophosphorylase

Table 3 Protein-protein interaction of selected drug targets of *Aspergillus terreus*

S. No.	Protein_id	Protein Name	Input	Number of interactions
1.	Q0CJ00	Protein rot1	ATEG_06334	10
2.	Q0CZH8	Homocitrate synthase	ATEG_00906	10
3.	Q0CSS2	Anthranilate synthase component I	ATEG_03262	10
4.	Q0CH64	Imidazoleglycerol-phosphate	ATEG_06978	10
5.	Q0C9E0	Glutamine amidotransferase subunit pdxT	ATEG_09694	4
6.	Q0CGU1	Biotin synthase dehydratase	ATEG_07101	10
7.	Q0CZ04	Dihydroxy-acid dehydratase	ATEG_01080	10
8.	Q0CSX2	Cyanide hydratase	ATEG_03213	1
9.	Q0C8V5	Thiamine-phosphate pyrophosphorylase	ATEG_09879	10

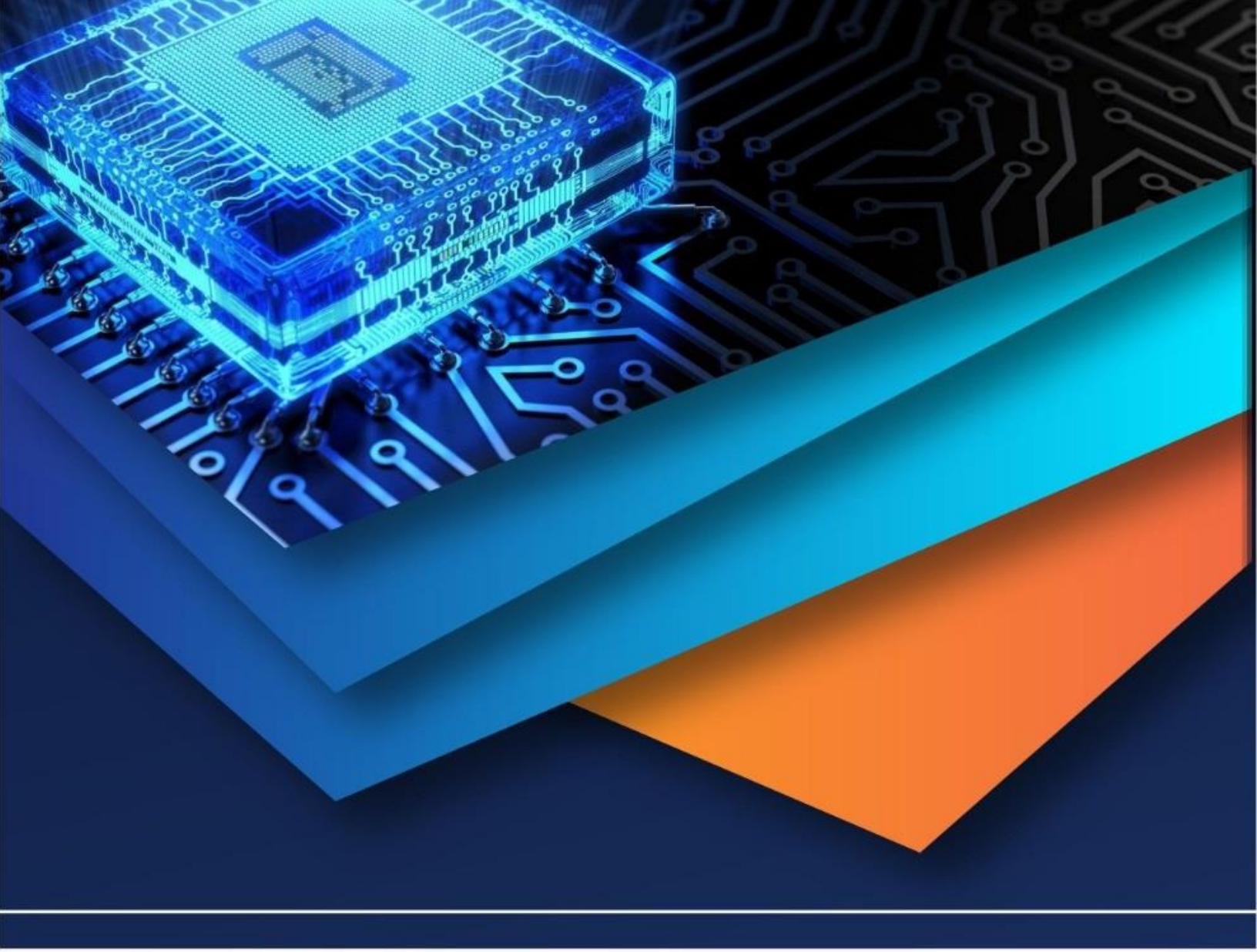
IV. CONCLUSION

Identification and characterization of probable drug targets with full genomic and proteomic sequences generated from the sequencing projects along with the computer-aided softwares is a new emerging trend in target based drug discovery. The main idea behind this study was to explore the knowledge about proteins responsible for virulence and pathogenicity in the organisms considered. The present study not only helped to identify a list of targets essential for the survival of pathogen but also it includes the targets which are less or non-toxic to host and organisms of gut flora. The current study helps in the characterization of the potential proteins that could be targets for efficient drug design against *Aspergillus terreus.* As subtractive proteomic approach is applied for the identification of drug targets, so the drug would be specific for the pathogen and not lethal to the host. Molecular modeling of the targets will decipher the best possible active sites that can be targeted by simulations for ligand based drug design.

Virtual screening against these potential targets followed by MD simulation might be useful in the discovery of potential therapeutic compounds against these fungal pathogens.

REFERENCES

[1] Tuli SS, "Fungal keratitis," Clinical Ophthalmology, vol.5, 275–279, 2011
[2] Kozakiewicz Z, "Aspergillus species on stored products," Mycological Papers, vol.161, 1–188, 1989.
[3] Bonnarme P, Gillet B, Sepulchre AM et al., "Itaconate biosynthesis in Aspergillus terreus," J. Bacteriol, vol. 177(12), 3573–3578, 1995
[4] Elinbaum S, Ferreyra H, Ellenrieder G et al., "Production of Aspergillus terreus alpha-L-rhamnosidase by solid state fermentation," Lett. Appl. Microbiol, vol. 34(1), 67–71, 2002.
[5] Wishart DS, "Drug-target discovery in silico: using the web to identify novel molecular targets for drug action," SEB Exp Biol Ser, vol. 58,145-76, 2007.
[6] Abadio AKR, Kioshima ES, Teixeira MM et al., "Comparative genomics allowed the identification of drug targets against human fungal pathogens," BMC Genomics, vol. 12(1), 75, 2011.
[7] Shinde P, Patil K, Malavkar N et al., "In Silico Subtractive Proteomics Approach for Identification and Characterization of Drug Targets in Streptococcus suis," Research and Reviews: Journal of Computational Biology, vol. 2(2),2319- 3433, 2013.
[8] KR, Sakharkar MK, and Chow VT, "A novel genomics approach for the identification of drug targets in pathogens, with special reference to Pseudomonas aeruginosa," In Silico Biol, vol. 4(3), 355–360, 2004.
[9] Dutta A, Singh SK, Ghosh P et al., "In silico identification of potential therapeutic targets in the human pathogen Helicobacter pylori," In Silico Biol, vol. 6(1-2), 43-47, 2006.
[10] Kumar S, Chaudhary K, Foster JM et al., "Mining Predicted Essential Genes of Brugia malayi for Nematode Drug Targets", PLoS ONE, vol. 2(11), e1189. 2007.
[11] Moolhuijzen PM, Lew-Tabor, AE, Wlodek BM et al., "Genomic analysis of Campylobacter fetus subspecies: identification of candidate virulence determinants and diagnostic assay targets," BMC Microbiol, vol. 9, 86, 2009.
[12] Amineni U, Pradhan D, Marisetty H, "In silico identification of common putative drug targets in Leptospira interrogans," J Chem Biol, vol. 3(4), 165-173, 2010.
[13] Butt AM, Nasrullah I, Tahir Set al., "Comparative genomics analysis of mycobacterium ulcerans for the identification of putative essential genes and therapeutic candidates," PLoS ONE, vol. 7(8), e43080, 2017.
[14] Huang Y, Niu B, Gao Y et al., "CD-HIT Suite: a web server for clustering and comparing biological sequences," Bioinformatics, vol. 26(5),680–682, 2010.
[15] Kerfeld CA, Scotts KM, "Using blast to reach E-value-tionary concepts," Plos Biol, vol. 9(2), 1-4, 2011
[16] McGinnis S, adden TL, "BLAST: at the core of a powerful and diverse set of sequence analysis tools," Nucleic Acids Research, vol. 32(2), 20–25, 2004.
[17] Zhang R, Lin Y, "DEG 5.0, a database of essential genes in both prokaryotes and eukaryotes," Nucleic Acids Research, vol. 37(Database issue),455–458, 2009
[18] Yu CS, Chen YC, Lu Chet al., "Prediction of protein subcellular localization." Proteins: Structure, Function and Bioinformatics, vol. 64(3),643-651, 2006
[19] Cai CZ, Han LY, Ji ZL et al., "SVM-Prot: web-based support vector machine software for functional classification of a protein from its primary sequence," Nucleic Acids Research, vol. 31(13),3692–3697, 2003
[20] Kanehisa M, Goto S, "KEGG: Kyoto encyclopedia of genes and genomes," Nucleic Acid Res, vol. 28(1), 27-30, 2000
[21] Moriya Y, Itoh M, Okuda S et al., "KAAS: an automatic genome annotation and pathway reconstruction server," Nucleic Acids Research, vol. 35(2), 182–185, 2007.
[22] Wishart DS, Knox C, Guo A C et al., "DrugBank: a comprehensive resource for in silico drug discovery and exploration," Nucleic Acid Research, vol. 34(suppl_1), 668–672, 2006. [23] Chen X, Ji ZL, Chen, Y.Z, "TTD:Therapeutic Target Database," Nucleic acid research, vol. 30(1),412-415, 2002
[23] Gao Z, Li H, Zhang H et al., "PDTD: a web-accessible protein database for drug target identification," BMC Bioinformatics, vol. 9(1),104, 2008
[24] Ye H, Ye L, Kang H et al., "HIT: linking herbal active ingredients to targets," Nucleic Acids Research, vol. 39 (suppl_1), 1055–1059, 2011
[25] Szklarczyk D, Franceschini A, Kuhn M et al., "The STRING database in 2011: functional interaction networks of proteins globally integrated and scored," Nucleic Acids Res, vol. 39(suppl_1), 561–568, 2011
[26] Shanmugham B, Pan A, "Identification and Characterization of Potential Therapeutic Candidates in Emerging Human Pathogen Mycobacterium abscessus: A Novel Hierarchical In Silico Approach," PLoS ONE, vol. 8(3), e59126, 2013
[27] Ludin P, Woodcroft B, Ralph SA et al., "In silico prediction of antimalarial drug target candidates," International Journal for Parasitology: Drugs and Drug Resistance, vol. 2, 191–199, 2012
[28] L J, Kuhn M, Stark M et al., "STRING 8–a global view on proteins and their functional interactions in 630 organisms," Nucleic Acids Res, vol. 37 (suppl_1), D412–D416, 2009.
[29] Kushwaha SK, Shakya M, "Protein interaction network analysis approach for potential drug target identification in Mycobacterium tuberculosis," J Theor Biol, vol. 262(2), 284–294, 2010.
[30] Cui T, Zhang L, Wang X et al., "Uncovering new signaling proteins and potential drug targets through the interactome analysis of Mycobacterium tuberculosis," BMC Genomics, vol. 10(1), 118, 2009.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY