Original Articles
Toxicology International, *Sept-Dec 2015 / Vol-22 / Issue-3, 59-69*
ISSN (Print) : 0971-6580
ISSN (Online) : 0976-5131
DOI: 10.22506/ti/2015/v22/i3/137626

Molecular Characterization and Drug Designing against Hypothetical Proteins of *Aspergillus fumigatus Af293* by Homology Modeling and Molecular Docking

Shah Md. Shahik[1], Towhidur Rahman[2], Srabasti Ghosh[2], Ismot Ara[3,4], Sabrina Amin[2], Md. Amzad Hossain[1], Yeasmeen Ali[1], Abdul Musaweer Habib[1] and Mohd. Omar Faruk Sikder[1*]

[1]Department of Genetic Engineering and Biotechnology, Faculty of Biological Sciences, University of Chittagong, Chittagong - 4331, Bangladesh; sikdergeb@cu.ac.bd
[2]Department of Biochemistry and Biotechnology, University of Science and Technology, Chittagong, Bangladesh
[3]Department of Computer Science and Engineering, Faculty of Science and Technology, Atish Dipankar University of Science and Technology, Dhaka - 1213, Bangladesh
[4]Department of Computer Science and Engineering, American International University, Dhaka -1213, Bangladesh

Abstract

Genome sequencing projects have led to an explosion of the huge amount of gene products in which many are of Hypothetical Proteins (HPs) with an unidentified function. Exploring and annotating the functions of HPs is important in *Aspergillus fumigatus* which is a pathogenic fungus that cause disease in individuals associated with immunodeficiency. In this study, sequence of five HPs of *A. fumigatus Af293,* which are nonhomologous to human, has been annotated from NCBI. Various computational tools such as three-dimensional (3D) structure was determined by means of homology modeling through Phyre2 and refined by ModRefiner. Then, the designed structure was evaluated with a structure validation program for instance ERRAT, PROCHECK, and QMEAN, for further structural analysis. The secondary structural features of proteins were determined through self-optimized prediction method with alignment and interacting networks by STRING. The receptor was analyzed by the active site and pocket finder tools. Docking studies were done through AutodockVina software. The 3D structure of five proteins was modeled and their ligand binding sites were identified. We have found domains and families of two proteins and these revealed that these proteins might have adenosine diphosphate-ribosylation activity, molecular chaperone, heat shock protein activity, plasminogen/hepatocyte growth factor activity, etc. Protein-ligand study of our predicted ligand shows the lowest energy of -7.4 kcal/mol for gi|66851587 and -6.7 kcal/mol for gi|66845370. Structural prediction of these proteins, detection of binding sites, and drug designing of current study will cover the way for further extensive investigation of these proteins in wet laboratory experiments and in that way assist drug design against Aspergillosis.

Keywords: Aspergillosis, *Aspergillus fumigatus Af293,* Docking Analysis, Homology Modeling, Hypothetical Protein

Received: 24-06-2015 **Revised:** 20-08-2015 **Accepted:** 22-08-2015

1. Introduction

Aspergillus fumigatus Af293, is a fungus of the genus Aspergillus. It is one of the most common Aspergillus species which cause disease in individuals associated with immunodeficiency. *A. fumigatus* is a saprophytic fungus in nature, is typically found in decaying organic matter and soil, such as compost heaps. Although it is

* *Author for correspondence* Email: sikdergeb@cu.ac.bd

not the most common fungus but most ubiquitous of those are with airborne conidia. The released conidia into the atmosphere have a diameter 2-3 μm to reach the lung Alveoli. According to environmental survey, all humans inhale several hundred of *A. fumigatus* conidia per day (Goodley *et al.,* 1994). Immunocompetent individuals rarely shows any adverse effect when inhale conidia, since the conidia are quickly eliminated by innate immune mechanism. The fungus is more likely to become pathogenic in immunocompromised individuals in developed countries (Dixon *et al.,* 1996). Invasive Aspergillosis (IA) is responsible for opportunistic fungal infections in patients with cancer, organ transplant recipients, and people with AIDS or leukemia (Abad *et al.,* 2010). At leukemia treatment centers, solid organ transplantation units and bone marrow transplantation are mentioned that IA is a major opportunistic organism which may cause death (Patel and Paya, 1997). *A. fumigatus* laboratory strain Af293 is isolated from a lung biopsy from a neutropenic patient with human lung invasive, which has a 29.4 MB genome consisting of eight chromosomes with 9926 predicted genes. Microarray analysis shows that expression of distinct set of genes are heat dependent and 700 genes are absent or importantly diverged in the closely related sexual species *Neosartoya fischeri*, which may have roles in the pathogenicity. There is a growing want for the automated annotation of proteins of unknown function, termed as "Hypothetical Proteins (HPs)" (Lubec *et al.,* 2005), the structures of that are far-famed. Structural genomics initiatives give full structures of HPs. However without performing function annotation, this structural goldmine is of very little use to biologists who have an interest specifically molecular system. The structure of many HPs is resolved in pipelines at structural genomics centers that typically lack the resources to interact in thorough functional characterization of the resolved structures (Islam *et al.,* 2015). Moreover, a number of the proteins that are thought to be well annotated might have extra functions on the far side their listed records. About 50%, the proteins in most genomes are candidates for HPs (Minion *et al.,* 2004). This cluster is of utmost importance to complete genomic and proteomic data. Detection of new HPs not only offers presentation of new structures however, conjointly new functions. A series of additional proteins pathway and cascades are going to be disclosed, finishing our fractional information on the mosaic of proteins intrinsically. The network of Protein-Protein Interactions (PPI) are going to be increasing logarithmically. New HPs are also serving as markers and pharmacologic targets. Detection of HP would be profitable to genomics sanctioning the invention of this point unknown or even predicted gene. HPs are a protein that is expected to be expressed from associate open reading frame, except for that there is no experimental proof of translation. With the overall belief that the majority of HPs are the product of pseudogenes, it is essential to possess a tool with the ability of pinpointing the minority of HPs with a high likelihood of being expressed (Desler *et al.,* 2009). There is no so far classification of HPs. HPs are predicted from nucleic acid sequences which have not been shown to exist by experimental protein chemical proof.

The lead compounds were then docked with HPs of *A. fumigatus* to study the interaction of inhibitors with the protein. These structures might represent up to half the potential coding regions of a genome.

2. Materials and Methods

2.1 Sequence Retrieval

HPs sequences of A. fumigatus were retrieved from NCBI (http://www.ncbi.nlm.nih.gov/) and finally nonhomologous to human, five HPs sequences gi|66850447, gi|66853579, gi|66851587, gi|66845370, and gi|71000381 were picked by the help of Basic Local Alignment Search Tool (http://blast.ncbi.nlm.nih.gov/Blast.cgi) (BLAST) analysis.

2.2 Physico-Chemical Characterization

Different properties including number of amino acids, amino acids composition (%), molecular weight, theoretical isoelectric point (pI), total number of positively (Arg + Lys), and negatively charged (Asp + Glu) residues, extinction co-efficient (Gill and von Hippel, 1989) instability index (Guruprasad *et al.,* 1990), Aliphatic Index (AI) (Ikai,1980), and grand average of hydropathicity (GRAVY) (Kyte and Doolittle, 1982) were computed by using the Expasy's Protparam tool (http://us.expasy.org/tools/protparam).

2.3 Protein Domain Prediction

Pfam (http://pfam.sanger.ac.uk/) is a collection of protein families; each represented by multiple protein sequence

alignments and Hidden Markov models, as well as it provides a good repository of models for identifying protein families, domain, and repeats (Finn *et al.,* 2006; Bru *et al.,* 2005). There are two components to Pfam database: Pfam-A and Pfam-B. Pfam-A, a set of high quality, manually curated and annotated models whereas Pfam-B has higher coverage but is fully automated (without manual curation).

2.4 Conserved Domain Database-Basic Local Alignment Search Tool

CD-Search is NCBI's interface which usually used for searching the conserved domain database with protein sequences. To quickly scan a set of pre-calculated position-specific scoring matrices with a protein query, it uses reversed position specific-BLAST and a variant of position-specific iterative-BLAST (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi/) (Marchler-Bauer *et al.,* 2011).

2.5 Motif Scan

Finding all known motifs that occur in a sequence is termed as motif scan. This allows one to paste a sequence and select the collections of motifs to scan for, and launch the search. A document page deals with the interpretation of the match scores (http://myhits.isb-sib.ch/cgi-bin/motif_scan).

2.6 Protein Localization Prediction

Usually, the protein's function is related to its subcellular localization. To predict subcellular localization directly from protein sequences useful for inferring protein functions. CELLO (http://cello.life.nctu.edu.tw/) and SOSUI (http://bp.nuap.nagoya-u.ac.jp/sosui/sosui_submit. html) server was used to characterize whether the protein is soluble or transmembrane in nature.

2.7 Stability of Proteins

DISULFIND (http://disulfind.dsi.unifi.it/) is a server which predicts the disulfide state of cysteines and their disulfide connectivity starting from sequence alone. Disulfide connectivity can be predicted from sequence and a bonding state assignment given as input and the result is a simple visualization of the assigned bonding state with confidence degrees and the most likely connectivity patterns, it may help the experimental structure determination process and may be useful in other genomic annotation tasks (Ceroni *et al.,* 2006).

2.8 Secondary Structure Prediction

The Self-Optimized Prediction Method with Alignment (SOPMA) has been described to improve the success rate in the prediction of the secondary structure of proteins. The input protein sequences were given in FASTA format. The conformational states were adjusted to four in order to predict helix, sheet, coil, and turn while other parameters were set as default.

2.9 Protein Three-Dimensional Structure Prediction

The three-dimensional (3D) structure of HPs was predicted by Phyre2 (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index), which accepts sequences in FASTA format.

2.10 Three-Dimensional Structure Refinement

ModRefiner (http://zhanglab.ccmb. med.umich.edu/ModRefiner/) used to construct and refine protein structures from C-alpha traces. Main-chain structures are first constructed from initial C-alpha traces and the side-chain is then refined together with the backbone atoms with the use of a composite physics and knowledge-based force field (Behjati *et al.,* 2012; Herráez, 2006). ModRefiner shows improvements in both global and local structures, which have better hydrogen-bonding networks, more accurate side-chain positions, and fewer atomic overlaps.

2.11 Three-Dimensional Structure Validation

The steriochemical quality of the models was verified with the program RAMPAGE (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php) and PROCHECK in order to select the best model (Xu and Zhang, 2011). The 3D-profilling of the residue was done by WHAT_CHECK, ERRAT, and VERIFY_3D (http://nihserver.mbi.ucla.edu/SAVES_3/).

2.12 Protein Function Prediction

Kiharalab (http://kiharalab.org/web/pfp.php) provides

functional analysis of proteins by classifying them into families and predicting domains and important sites. SMART (http://dag.embl-heidelberg.de/hnb_cgi/show_input_page.pl?taskId=_notask&UserId=_nouser) reports on the functional domains and features of desired proteins. ProFunc (http://www.ebi.ac.uk/thornton-srv/databases/profunc/index.html) helps to identify the likely biochemical function of a protein from its 3D structure.

2.13 Protein-Protein Interactions

STRING (http://string.embl.de/) is a database of known and predicted protein interactions (Szklarczyk *et al.,* 2011). The interactions include direct (physical) and indirect (functional) associations; they are derived from four sources: Genomic context, conserved co-expression, high-throughput experiments, and previous knowledge. STRING quantitatively integrates interactions data from these sources for a large number of organisms and finally transfers information between these organisms where applicable.

2.14 Prediction of Active Site

CASTpisan automated software was used to predict the active site with the determined model. Analysis of active sites is vital task for the modeled protein to do further work on its docking studies and it gives us an idea to make a grid before docking (Sibi and Paul, 2013; Jambon, 2005).

2.15 Molecular Docking Study

Molecular docking is a computational procedure to calculate the conformation of a receptor-ligand complex. It is used to recognize the binding affinity and interaction energy of the molecules with the target protein. Auto Dock Vina (Trott and Olson, 2010) was performed to analysis docking. Before starting the docking stimulations, HPs of *A. fumigatus* was modified by adding polar hydrogen, eliminating all the water molecules, and the grid box was set for its binding site, whereas all the torsion angle of ligands were set free by the ligand module. Binding energy was evaluated by 3D grid box (box size, 100 Å × 100 Å × 100 Å; box center, −2.487 × 27.904 × 75.494 for x, y, and z, respectively) for gi|66851587 and grid box (box size, 90 Å × 100 Å × 100 Å; box center, 20.384 × 51.353 × −7.097 for x, y, and z, respectively) for gi|66845370, was created using Auto Grid, which calculates the grid map. The combined binding with target HPs *A. fumigatus Af293* and ligand, Neferine was obtained by using PyMOL (Patel *et al.,* 2014).

3. Result and Discussion

The physiochemical properties of these HPs of *A. fumigatus* are analyzed for the first time. In Table 1 the physicochemical properties of HPs are given. Isoelectric point (pI) of the HPs arrays from 5.94 to 9.00. In the pH level, at which the amino acid of protein bears no net charge and hence does not move in a direct electrical field is cold isoelectric point (pI). Moreover, proteins come to be stable and compact at isoelectric pH, for this purpose computed isoelectric point (pI) will be helpful for developing a buffer system for purification by isoelectric focusing method. The extinction co-efficient of HPs ranges from 24075 to 94225 M cm calculated by Expasy's Protparam instead of 276, 278, 279, and 282 nm. The existence of high concentration of Cys, Trp, and Tyr indicates a higher extinction coefficient of the proteins. The quantitative study of protein-protein and protein-ligand interactions in solution can be done by using extinction coefficients. The instability index value of the HPs from 34.20 to 61.00 was found. It is predicted that a protein will be unstable whose instability index is >40 and value <40 will be stable. For structure identification of protein, instability index is another parameter. Proteins, gi|66853579 and gi|66845370 were stable and others were unstable. The instability index indicates an approximate stability of proteins. The AI is the relative volume of a protein engaged by aliphatic side chains (A, V, I and L) and is considered as a positive factor for the raise of thermal stability of globular proteins. The range of the AI is from 54.30 to 86.70. The proteins with very high AI may show stability in a wide-ranging temperature where lower AI proteins are not heat stable and show more flexibility. The grand average hydropathy (GRAVY) is ranging from (-0.232) to (-0.939). The better interaction of protein and water is occurring in low GRAVY. The GRAVY value for a protein is calculated by adding the values of hydropathy of all the amino acids and dividing it by the number of residues in the sequence (Ikai, 1980).

To study the functional analysis conserved domains

Table 1. Physicochemical properties of HPs of Protparam tool

Sequence ID	Number of AA	MW	pI	(+) R	(−) R	EC	II	AI	GRAVIY
gi|66850447	348	38500.6	9.00	29	33	35535	61.00	59.83	−0.803
gi|66853579	487	55010.0	5.94	56	45	94225	36.61	84.29	−0.334
gi|66851587	330	37704.8	6.54	51	43	24075	41.93	54.30	−0.939
gi|66845370	215	24102.4	8.52	21	23	28545	34.20	80.28	−0.232
gi|71000381	443	50926.3	6.47	54	51	84380	48.86	86.70	−0.316

HPs=Hypothetical proteins, MW=Molecular weight, AI=Aliphatic index

were observed because it is the functional units within a protein that act as building blocks in molecular evolution and recombine in various arrangements to make proteins with different functions. The proteins have been classified into particular families based on matches to specific super families history, identification of proteins with specific similar domains, in the sequence (Marchler-Bauer *et al.,* 2011). In our study, we used 5 HPs possessing different specific domains but only 1 protein gi|66851587 contain multi domain. The presence of these domains in the HPs indicates that the protein might do the same function. Domains and families present in HPs were identified by the Pfam database research (Tables 2 and 3). They are GAT_SF, IbpA, ACD_sHsps-like, Thioredoxin_like, PAN_APPLE domain.

To explain the protein functions involved in various cellular processes, it is important to know the subcellular localization of that protein. During the drug discovery process knowledge of the subcellular localization of a protein play a very significant role in target identification. In our study, we have found one protein gi|71000381 was cytoplasmic as their best performing sites, gi|66853579 at plasma membrane, gi|66845370 in extracellular and remaining two proteins gi|66850447 and gi|66851587 in nuclear. The server SOSUI discriminates whether the HPs are soluble or membranous. All were soluble, and no transmembrane protein was found. Moreover, disulfide bridges DISULPHIND server play a major role in stabilizing the folding process of many proteins. Server revealed that no disulpfide bonds were present in any of those proteins which indicate that they were thermally unstable (Islam *et al.,* 2014). Disulfide bridges are very important finding in the study of structural and functional properties of specific proteins (Ceroni *et al.,* 2006).

For performing almost all the cellular functions, the PPI are important. Proteins often interact with one another in a mutually dependent way to perform a common function. It is notable that translational factors interact among themselves to carry out the whole translation. The function of a protein is predictable from this based on their interaction with other proteins. It is very rare that proteins bring out function with any interactions with other biomolecules. For this reason, in this post genomic era PPI databases have turned as a most important resource for searching biological networks and pathways in cells (Laurie and Jackson, 2005). The protein gi|66851587 had interacted with several proteins such as Hsp70 Nucleotide Exchange Factor (NEF) fes1; functions as a NEF for

Table 2. Functional description of super-families of HPs

Sequence ID	Domains	Description
gi|66850447	GAT_SF	GAT domain found in eukaryotic Arf-binding proteins (GGAs), metazoan myb protein 1 (Tom1) - like proteins, and similar proteins
gi|66851587	IbpA	Molecular chaperone (small heat shock protein) posttranslational modification, protein turn-over, chaperones
	ACD_sHsps-like	ACD found in alpha-crystallin-type small heat shock proteins, and a similar domain found in p23 (a cochaperone for Hsp90) and in other p23-like proteins. And (specific hit) COG0071, molecular chaperone (small heat shock protein)
gi|66845370	Thioredoxin_like	Protein disulfide oxidoreductases and other proteins with a thioredoxin fold
gi|71000381	PAN_APPLE	PAN/APPLE-like domain; present in N-terminal (N) domains of plasminogen/ hepatocyte growth factor proteins

HPs=Hypothetical proteins, Arf=ADP-ribosylation factor, ACD=Alpha-crystallin domain

Table 3. Identification of domains by CDD-BLAST

Sequence ID	Domains	Envelope	Alignment	HMM	HMM length	Bit score	*E* value
gi|66850447	GAT_SF	132–230	137–220	9–91	101	52.0	$5.2e^{-14}$
gi|66853579	-	-	-	-	-	-	-
gi|66851587	IbpA	187–297	190–287	4–93	102	22.2	$8.9e^{-05}$
	ACD_sHsps-like	077–281	187–281	-	95	-	-
gi|66845370	AhpC/TSA antioxidant enzyme	077–192	089–192	15–115	115	60.5	$1.3e^{-16}$
gi|71000381	PAN_APPLE	371–411	374–408	181–200	35	29.2	$4.7e^{-07}$

HMM: Hidden Markov models, CDD-BLAST=Conserved domain database-Basic Local Alignment Search Tool

Hsp70 chaperones which accelerates the release of adenosine diphosphate (ADP), required for fully efficient Hsp70-mediated folding of proteins (confidence score, 0.829), Afu5g07340 (Alt. Locus), DnaJ domain protein PSI (confidence score, 0.827), Afu1g15270 (Alt. Locus), heat shock protein Hsp98/Hsp104/ClpA (confidence score, 0.787), Afu1g11180 (Alt. Locus) heat shock protein/chaperonin HSP78, polyubiquitin fragment, heat shock protein Hsp88, Afu7g08320 (Alt. Locus) heat shock transcription factor, Afu5g01900 (Alt. Locus) heat shock transcription factor Hsf1, Afu5g10270 (Alt. Locus) heat shock protein Hsp20/Hsp26, and Afu1g05040 (Alt. Locus) protein mitochondrial targeting protein. gi|66845370 interacting with arginine N-methyltransferase 2, delta-N-monomethylargininetransferase (confidence score, 0.608). Other HPs do not interact with any other proteins. Figure 1 and Table 4 indicate the protein-protein interacting networks of HPs, which might have functions of their interacting proteins (Peri *et al.,* 2003).

Secondary structure analysis is developing day by day to predict protein function and structure. The secondary structures of HPs were predicted by SOPMA with standard parameters are presented in Table 5. By using this software, Alpha helix (Hh), 310 helix, (Gg), Pi helix (Ii), Beta bridge (Bb), Extended strand (Ee), Beta turn (Tt), Bend region (Ss), Random coil (Cc), and Ambigous states were predicted. Homology modeling is the most frequent structure prediction method for 3D structure. Target protein sequences were used to predict the 3D structure of the target protein using Protein Structure Prediction Server (PS) 2 (Figure 2). Z-score, ERRAT and Ramachandram plots were used to check the quality and reliability of the structure. PROCHECK checks the stereochemical quality of a protein structure by analyzing residue-by-residue geometry and overall structure geometry. This tool was used in determining the Ramachandran plot to assure the quality of the model. The consequences of the Ramachandran plots are tabulated (Table 6) and the reliability of the model was further tested by ERRAT, which evaluates the statistics of nonbonded interactions between different atom types and plots the value of the error function versus position of a 9-residue gliding window, calculated by an assessment with statistics from highly refined structures. The *Z*-scores, Ramachandran plot, and ERRAT results confirmed the quality of the homology model of the HPs. The functions of proteins under study were predicted by PFP from Kihara Bioinformatics Laboratory, which uses references from gene ontology. These are shown in Table 7.

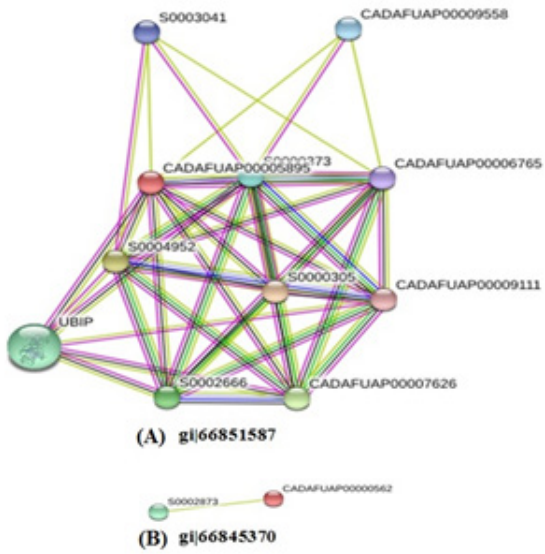


Figure 1. Protein-protein interactions of *Aspergillus fumigatus Af293*. (a) gi|66851587 had interacted with several proteins. (b) gi|66845370 interacting with only one protein.

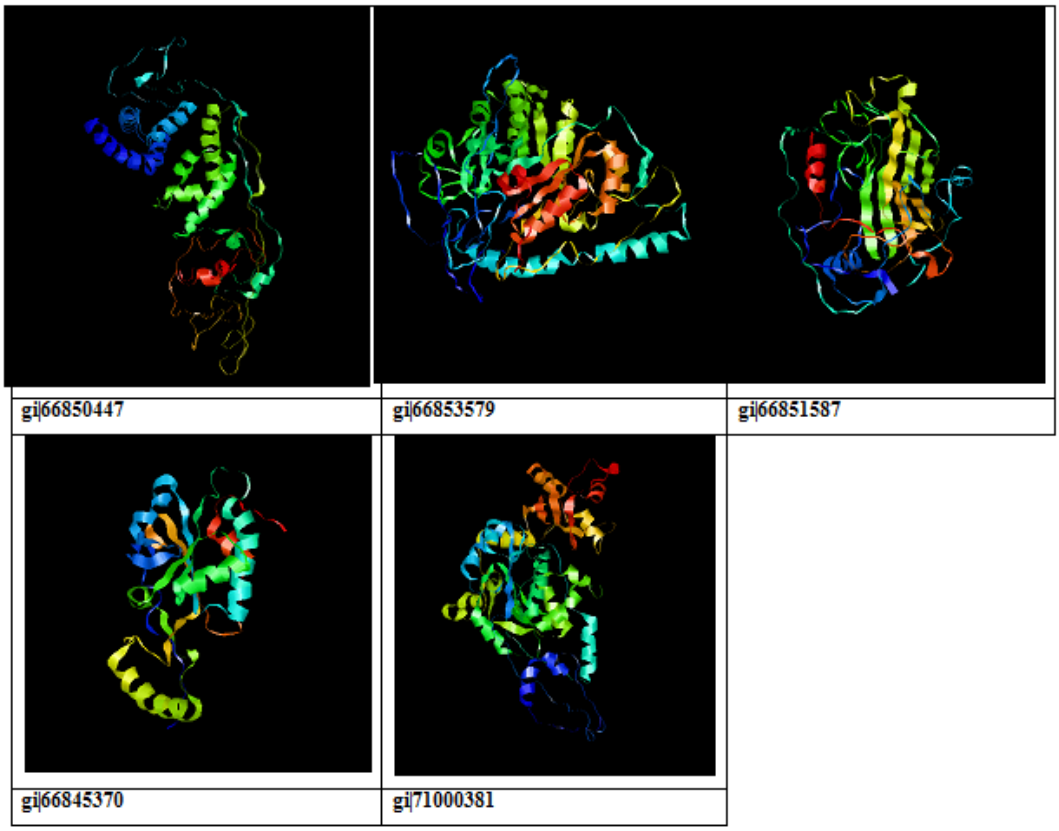


Figure 2. Three-dimensional of hypothetical proteins of *Aspergillus fumigatus Af293.*

The active site of HPs of *A. fumigatus*Af293 was predicted by using CASTp. Further, in this study, we have also reported the best active site area of the experimental protein, as well as the number of amino acids involved in it. Figure 3 shows the interacting residues, which was found in the active site of gi|66851587and gi|66845370.

Automated docking was used to evaluate the binding modes and conformation of the ligand molecule. Understanding the docking studies with HPs and neferine, analysis of the lowest docked energy value, the involvement of H-bonds, calculated root mean square deviation value, and interacting residues was considered (Table 8). The smallest amount of docked energy assumes a better docking result. Receptor-ligand study of our predicted ligand shows the lowest energy of -7.4 kcal/mol for gi|66851587 and -6.7 kcal/mol for gi|66845370, as well as a root mean square distance of 0.000Å, and it contains with hydrogen bonds. Neferine (4-[(6,7-dimethoxy-2-methyl-3,4-dihydro-1H-isoquinolin-1-yl)methyl]-2-[[6-methoxy-1-[(4-methoxy phenyl)methyl]-2-methyl-3,4-dihydro-1H-isoquinolin-7-yl]oxy]phenol), which has a molecular weight of 625.786 g/mol, rotatable bonds is 10 and net charge 1 (Table 9), was found to bind at the active site of HPs with the lowest binding energy. The interaction between the active site residues and the ligand found in this study is useful for perceiving the potential mechanism of residues and the drug binding. It has been clear that neferine formed similar hydrogen bond interactions with HPs. The hydrogen bonds play a significant role in the structure and function of biological molecules, and we have found significant results. These molecules were completely wrapped by active site amino acid residues at the active site pocket region (Figure 4). From the active site analysis, 115 amino acid residues for gi|66851587, and 10 amino acid residues for gi|66845370 were found in the potent active site. Docking analysis with ligand identified specific residues in the pocket region, which further itself inhibit its function and exposes studies about new drug discovery.

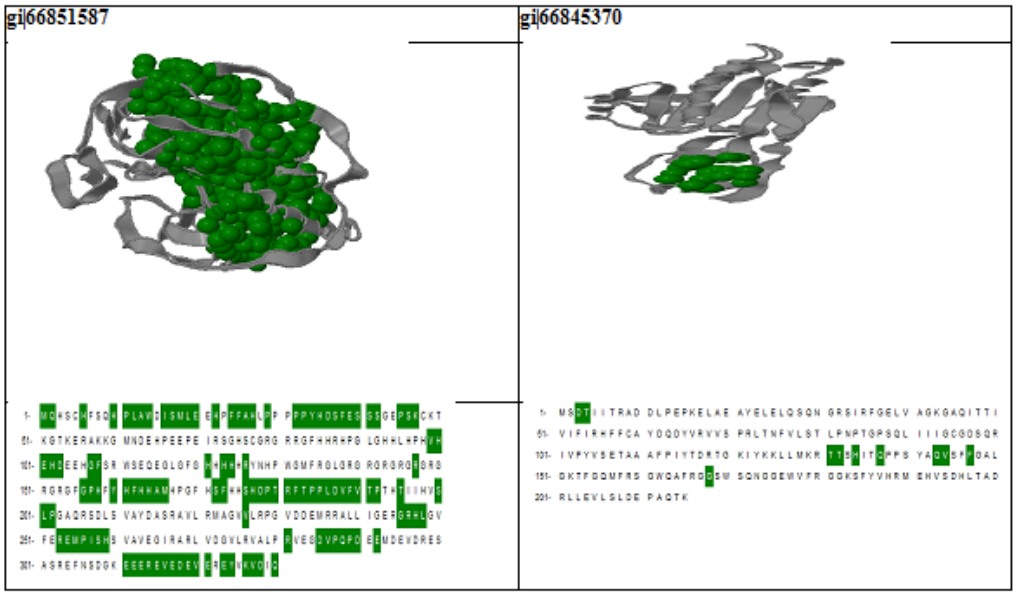


Figure 3. Pockets in the motif of hypothetical proteins of *Aspergillus fumigatus AF293* and largest active site. The amino acid residues in the active site green color predicted by CASTP.

Table 4. HPs interacting with functionally important proteins

Sequence ID	Interacting protein	Score
gi|66850447	Not found	-
gi|66853579	Not found	-
gi|66851587	Hsp70 NEF fes1	0.829
	Afu5g07340 (Alt. Locus), DnaJ domain protein PSI, putative, similar to PSI protein (*S. pombe*)	0.827
	Afu1g15270 (Alt. Locus), heat shock protein Hsp98/Hsp104/ClpA, putative, similar to heat shock protein ClpA (*P. brasiliensis*)	0.787
	Afu1g11180 (Alt. Locus), heat shock protein/chaperonin HSP78, putative, similar to heat shock protein 78	0.748
	Polyubiquitin fragment	0.683
	Heat shock protein Hsp88, putativeHsp70 chaperone Hsp88	0.628
	Afu7g08320 (Alt. Locus), heat shock transcription factor, putative, similar to related to heat shock transcription factor (*N. crassa*)	0.615
	Afu5g01900 (Alt. Locus), heat shock transcription factor Hsf1, putative, similar to trimeric heat shock transcription factor, activates multiple genes in response to hyperthermia; recognizes variable HSEs consisting of inverted NGAAN repeats; constitutively DNA-bound; posttranslationally regulated; Hsf1p (*S. cerevisiae*)	0.615
	Afu5g10270 (Alt. Locus), heat shock protein Hsp20/Hsp26, putative, similar to HspA protein (*S. aurantiaca*)	0.607
	Afu1g05040 (Alt. Locus), protein mitochondrial targeting protein (Mas1), putative, similar to mitochondrial protein import protein MAS5; Protein YDJ1 (*S. cerevisiae*)	0.608
gi|66845370	Arginine N-methyltransferase 2 (EC 2.1.1.-); Delta-N-monomethylargininetransferase. May monomethylate ribosomal protein L12 (by similarity)	0.608
gi|71000381	Not found	---

HSEs=Heat shock elements, *S. pombe=Schizosaccharomyces pombe*, *P. brasiliensis=Paracoccidioides brasiliensis*, *N. crassa=Neurospora crassa*, *S. cerevisiae=-Saccharomyces cerevisiae*,*S. aurantiaca=Stigmatella aurantiaca*, NEF=Nucleotide exchange factor

Table 5. Calculated secondary structure elements of HPs of *A. fumigatus Af293* by SOPMA

Sequence ID	Alpha helix	310 helix	Pi helix	Beta bridge	Extended strand	Beta turn	Bend region	Random coil	Ambiguous states
gi|66850447	36.21	0.00	0.00	0.00	6.03	1.15	0.00	56.61	0.00
gi|66853579	31.83	0.00	0.00	0.00	12.32	3.70	0.00	52.16	0.00
gi|66851587	25.45	0.00	0.00	0.00	13.94	9.70	0.00	50.91	0.00
gi|66845370	34.88	0.00	0.00	0.00	14.42	3.72	0.00	46.98	0.00
gi|71000381	38.37	0.00	0.00	0.00	13.54	4.51	0.00	43.57	0.00

SOPMA=Self-optimized prediction method with alignment, *A. fumigatus=Aspergillus fumigatus*

Table 6. Ramachandran plot statistics of HPs of *A. fumigatus Af293*

Sequence ID	Number of residues in favored region (%)	Number of residues in allowed region (%)	Number of residues in outlier region (%)
gi|66850447	95.1	3.8	1.2
gi|66853579	94.6	4.3	1.0
gi|66851587	95.4	4.0	0.6
gi|66845370	99.1	0.5	0.5
gi|71000381	94.6	4.8	0.7

HPs=Hypothetical proteins, *A. fumigatus=Aspergillus fumigatus*

Table 7. Function prediction of HPs of *A. fumigatus Af293*

Sequence ID	Molecular function	Biological process	Cellular component
gi|66850447	Protein binding, translation elongation factor activity, Translation factor activity, nucleic acid binding	Translation elongation, protein metabolic process, macromolecule metabolic process, translation, metabolic process	Intracellular, cell, cell part, intracellular part, cytoplasm, membrane-bound organelle
gi|66853579	Phosphoric diester hydrolase activity, phosphoric ester hydrolase activity	Lipid metabolic process	Cell part, cell, cytoplasm, intracellular, intracellular part membrane
gi|66851587	Ion binding, cation binding, metal ion binding, structural constituents of eye lens, binding, unfolded protein binding	Response to stress, stimulus, heat, temperature stimulus, hydrogen peroxide, high light intensity	Cytoplasm, intracellular, intracellular part, cell, cell part, nucleus
gi|66845370	Nucleic acid binding, binding, ATP binding, helicase activity, RNA binding, protein ribonucleotide binding	Processing nucleic acid metabolic, RNA metabolic process, DNA metabolic, metabolic process, viral reproduction	Virion, membrane, viral envelope, cell, cell part, nucleus
gi|71000381	Transferase activity, metal ion binding, cation binding, galactosyltransferase activity	Macromolecule and protein glycosylation, glycoprotein biosynthetic process, angiogenesis	Membrane, integral to membrane, intrinsic to membrane, membrane part, cell part, cell, golgi membrane

HPs=Hypothetical proteins, *A. fumigatus=Aspergillus fumigatus*

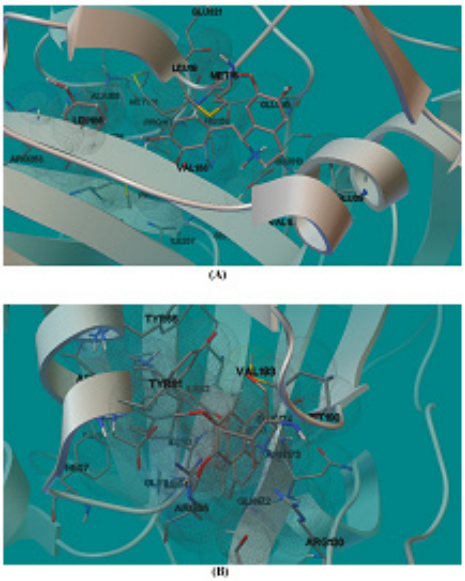


Figure 4. Binding mode of neferine within the active site of. (a) gi|66851587. (b) gi|66845370.

4. Conclusion

We need to discover the structural and functional annotation properties of HPs in the pathogenic fungus *A. fumigatus* AF293, which produce many virulence factors and cause severe infections and disease. We retrieved all HPs from NCBI database and finally pick five HPs on the basis of nonhomologous to human for drug designing. Characterization of its physicochemical properties is identified domains and families using various bioinformatics tools and databases. The refined structures were modeled and their ligand binding sites were identified. The analysis shown functionally important domains are families which were involved in ADP-ribosylation activity, molecular chaperone, heat shock protein activity, plasminogen/hepatocyte growth factor activity etc. The modeled 3D structure will provide a good-quality foundation for experimental development of the crystal structure, and molecular docking study will assist efficient inhibitor design against aspergillosis.

5. Acknowledgment

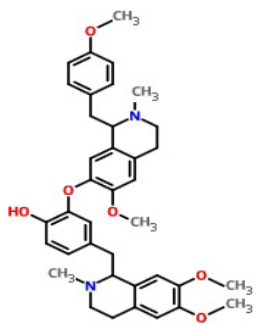
Authors would like to thank Md. Anayet Hasan (Department of Genetic Engineering and Biotechnology, University of Chittagong, Bangladesh) for his technical support for the docking studies.

Table 8. Comparative docking study of the ligand to the target

Protein	Ligand	Dock score (kcal/mol)	RMSD l.b. (Å)	RMSD u.b.(Å)	Interacting residues
gi|66851587	Neferine	−7.9	0.00	0.00	MET18, LEU19, GLU39, GLY156, PRO157, HIS158, ALA165, MET166, LEU186, VAL188, VAL190, ARG253, MET255, PRO256, ILE257, SER258, GLU313, GLU315, GLU321
gi|66845370	Neferine	−6.7	0.00	0.00	ILE52, PHE53, ILE54, ARG55, PHE57, TRY61, ASP62, TYR65, VAL66, ILE92, ILE93, GLY94, ILE114, ARG130, SER133, GLY172, ASN173, GLY174, MET190, VAL193

RMSD=Root mean square deviation, *A. fumigatus*=*Aspergillus fumigatus*

Table 9. Description of the ligand molecule

Name	Neferine	Chemical structure
Identifiers	4-[(6,7-dimethoxy-2-methyl-3,4-dihydro-1H-isoquinolin-1-yl)methyl]-2-[[6-methoxy-1-[(4-methoxyphenyl)methyl]-2-methyl-3,4-dihydro-1H-isoquinolin-7-yl]oxy]phenol	
Formula	$C_{38}H_{44}N_2O_6$	
Molecular weight (g/mol)	625.786	
pH (at mid-range)	6–8	
Rotatable bonds	10	
Net charge	1	

6. References

1. Goodley JM, Clayton YM, Hay RJ. Environmental sampling for *Aspergilli* during building construction on a hospital site. J Hosp Infect. 1994; 26:27–35.
2. Dixon DM, McNeil MM, Cohen ML, Gellin BG, La Montagne JR. Fungal infections: A growing threat. Public Health Rep. 1996; 111:226–35.
3. Abad A, Fernández-Molina JV, Bikandi J, Ramírez A, Margareto J, Sendino J, *et al.* What makes *Aspergillusfumigatus* a successful pathogen? Genes and molecules involved in invasive aspergillosis. Rev Iberoam Micol. 2010; 27:155–82.
4. Patel R, Paya CV. Infections in solid-organ transplant recipients. Clin Microbiol Rev 1997; 10:86–124.
5. Lubec G, Afjehi-Sadat L, Yang JW, John JP. Searching for hypothetical proteins: Theory and practice based upon original data and literature. Prog Neurobiol. 2005; 77:90–127.
6. Islam MS, Shahik SM, Sohel M, Patwary NI, Hasan MA. *In silico* structural and functional annotation of hypothetical proteins of *Vibrio cholerae* O139. Genomics Inform 2015; 13:53–9.
7. Minion FC, Lefkowitz EJ, Madsen ML, Cleary BJ, Swartzell SM, Mahairas GG. The genome sequence of *Mycoplasmahyopneumoniae* strain 232, the agent of swine mycoplasmosis. J Bacteriol. 2004; 186:7123–33.
8. Desler C, Suravajhala P, Sanderhoff M, Rasmussen M, Rasmussen LJ. *In silico* screening for functional candidates amongst hypothetical proteins. BMC Bioinformatics 2009; 10:289.
9. Gill SC, von Hippel PH. Calculation of protein extinction coefficients from amino acid sequence data. Anal Biochem 1989; 182:319–26.
10. Guruprasad K, Reddy BV, Pandit MW. Correlation between stability of a protein and its dipeptide composition: A novel approach for predicting *in vivo* stability of a protein from its primary sequence. Protein Eng. 1990; 4:155–61.
11. Ikai A. Thermostability and aliphatic index of globular proteins. J Biochem. 1980; 88:1895–8.
12. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. J MolBiol. 1982; 157:105–32.
13. Finn RD, Mistry J, Schuster-Böckler B, Griffiths-Jones S, Hollich V, Lassmann T, *et al.* Pfam: Clans, web tools and services. Nucleic Acids Res. 2006; 34:D247–51.
14. Bru C, Courcelle E, Carrère S, Beausse Y, Dalmar S, Kahn D. The ProDom database of protein domain families: More emphasis on 3D. Nucleic Acids Res. 2005; 33:D212–5.
15. Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, *et al.* CDD: A conserved domain database for the functional annotation of proteins. Nucleic Acids Res. 2011; 39:D225–9.
16. Ceroni A, Passerini A, Vullo A, Frasconi P. DISULFIND: A disulfide bonding state and cysteine connectivity prediction server. Nucleic Acids Res. 2006; 34:W177–81.
17. Behjati M, Torktaz I, Mohammadpour M, Ahmadian G, Easton AJ. Comparative modeling of CCRL1, a key protein in masked immune diseases and virtual screening for finding inhibitor of this protein. Bioinformation. 2012; 8:336–40.
18. Herráez A. Biomolecules in the computer: Jmol to the rescue. Biochem Mol Biol Educ 2006; 34:255–61.
19. Xu D, Zhang Y. Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. Biophys J. 2011; 101:2525–34.
20. Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, *et al.* The string database in 2011: Functional interaction networks of proteins, globally integrated and scored. Nucleic Acids Res. 2011; 39:D561–8.
21. Sibi P, Paul M. *In silico* docking analysis of constituents of *Zingiberofficinale* as antidepressant. J Pharmacogn Phytother. 2013; 5:101–5.
22. Jambon M, Andrieu O, Combet C, Deléage G, Delfaud F, Geourjon C. The SuMo server: 3D search for protein functional sites. Bioinformatics. 2005; 21:3929–30.
23. Trott O, Olson AJ. AutoDockVina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem. 2010; 31:455–61.
24. Patel H, Grüning BA, Günther S, Merfort I. PyWATER: APyMOL plug-in to find conserved water molecules in proteins by clustering. Bioinformatics. 2014; 30:2978–80.
25. Islam MS, Patwary NI, Muzahid NH, Shahik SM, Sohel M, Hasan MA. A systematic study on structure and function of ATPase of *Wuchereriabancrofti*. ToxicolInt. 2014; 21:269–74.
26. Laurie AT, Jackson RM. Q-SiteFinder: An energy-based method for the prediction of protein-ligand binding sites. Bioinformatics. 2005; 21:1908–16.
27. Peri S, Navarro JD, Amanchy R, Kristiansen TZ, Jonnalagadda CK, Surendranath V, *et al.* Development of human protein reference database as an initial platform for approaching systems biology in humans. Genome Res. 2003; 13:2363–71.

Source of Support: Nil. **Conflicts of Interest:** None declared.