Volume 15 • Number 3 • September 2024 • em00837

JOURNAL OF CLINICAL AND EXPERIMENTAL INVESTIGATIONS

RESEARCH ARTICLE

Computational characterization and structure-driven functional exploration of a hypothetical protein from *Candida auris*

Nur Alam [1], *Tanvir Hossain Emon* [1], *Md. Anwar Hossain* [1], *Md. Abunasar Miah* [1], *Farzana Akter* [1], *Md. Abdul Kader* [1], *Md. Nur Islam* [1], *Md. Easin Mia* [1*]

[1] Department of Biotechnology and Genetic Engineering, Noakhali Science and Technology University, Noakhali-3814, BANGLADESH

ABSTRACT

Candida auris, a fungal species, has emerged as a global menace due to its drug-resistant nature, leading to widespread invasive infections. Currently, there is no vaccine to prevent *C. auris*. The study was attempted to ascertain the structure and role of an unannotated hypothetical protein (HP) (accession no. QWW22972.1) from *C. auris* utilizing various bioinformatics tools. In this study, HP was found to be stable and polar, located in the cytoplasm. Various tools like NCBI-CD search, ScanProsite, InterPro, and SMART, identified it as a member of the Ran family of GTP-binding nuclear proteins that involves facilitating nucleocytoplasmic transport, including the import and export of proteins and RNAs during the interphase of mitosis. The protein's secondary structure analysis indicated a dominance of the alpha helix. Its three-dimensional (3D) structure, modeled via the SWISS-MODEL server using a template protein with a 94.15% sequence identity, was validated by PROCHECK, QMEAN, Verify3D, and ERRAT tools. After YASARA energy reduction, a more stable 3D structure was visible. Furthermore, protein-protein interactions were obtained from STRING server, and active site were derived from the computed atlas of surface topography of proteins server. However, this study may enhance understanding of the molecular foundation of the HP and help identify potential therapeutic targets.

Keywords: *Candida auris*, bioinformatics tools, functional annotation, hypothetical protein, *in silico* characterization, homology modeling

Correspondence:

Md. Easin Mia

Address: Department of Biotechnology and Genetic Engineering, Noakhali Science and Technology University, Noakhali-3814, BANGLADESH

Email: easin.bge@nstu.edu.bd

Received: 02.11.2023,
Accepted: 08.08.2024
https://doi.org/10.29333/jcei/15040

INTRODUCTION

Candida auris is a fungus that causes candidiasis in human and is becoming increasingly resistant to antifungal treatments. Its high level of contagiousness and the challenges associated with its treatment make it a significant public health concern [1-3]. Cases of *C. auris* have been found in intensive care units and other healthcare facilities worldwide [4, 5]. The first instance of *C. auris* was detected in a person's ear canal in Japan [6]. It was linked to bloodstream infections in South Korea in 2011, and subsequently, it has been reported in more than 60 other countries [7]. It is challenging to treat *C. auris* since an increasing proportion of newly reported cases have been shown resistant to common antifungal medications such as azoles, echinocandins, and polyenes [8]. The worst thing is that *C. auris* has been discovered a multidrug-resistant fungus that has infected people in more than 30 nations, including the United States [9]. *C. auris* frequently establishes itself on the skin, in the respiratory system, and within the urinary tract. Additionally, it can contaminate surfaces and equipment via fluids released from the skin. In hospital settings, infection transmission can occur among individuals through direct or indirect contact [2,10].

Urgent collection of huge information about *C. auris* is imperative for identifying an effective treatment. Proteins with concealed functions or those that have not yet undergone experimental characterization denoted as hypothetical proteins (HPs). They are anticipated to be

Table 1. Tools employed for the computational characterization of the protein

Function	Tools/Servers	URL	Reference
Sequence retrieval	NCBI	https://www.ncbi.nlm.nih.gov/	[15]
Physiochemical properties analysis	ExPASy ProtParam	https://web.expasy.org/protparam/	[17]
Subcellular localization	PSORT Ⅱ	https://psort.hgc.jp/form2.html	[18]
	WoIF PSORT	https://wolfpsort.hgc.jp/	[19]
	Euk-mPLoc 2.0	http://www.csbio.sjtu.edu.cn/bioinf/euk-multi-2/	[20]
Functional characterization	Conserved domain database	https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi	[21]
	ScanProsite	https://prosite.expasy.org/scanprosite/	[22]
	InterPro	http://www.ebi.ac.uk/interpro/	[23]
	SMART	http://smart.embl-heidelberg.de/	[24]
	PFP-FunD SeqE	http://www.csbio.sjtu.edu.cn/bioinf/PFP-FunDSeqE/	[25]
Sequence similarity search	BLASTp	https://blast.ncbi.nlm.nih.gov/Blast.cgi	[26]
Secondary structure prediction	PSIPRED	http://bioinf.cs.ucl.ac.uk/psipred/	[27]
	SOPMA	https://npsaprabi.ibcp.fr/cgibin/npsa_automat.pl?page=/NPSA/npsa_sopma.html	[28]
3D Structure prediction	SWISS-MODEL	http://swissmodel.expasy.org/	[29]
3D Structure validation	PROCHECK, Verify3D, & ERRAT	https://saves.mbi.ucla.edu/	[30-32]
	QMEAN4	https://swissmodel.expasy.org/qmean/	[33]
	ProSA-web (Z scores)	https://prosa.services.came.sbg.ac.at/prosa.php	[34]
Protein-protein interaction analysis	STRING	https://string-db.org/	[35]
Active site detection	CASTp	http://sts.bioe.uic.edu/castp/	[36]
Energy minimization of 3D structure	YASARA	http://www.yasara.org/minimizationserver.htm	[37]

present because of the resemblance in their sequence to other proteins, but it is unclear how they are capable of doing. However, HPs have an essential component of the proteome and are present in all organisms [11]. In our investigation, we employed multiple bioinformatics tools with diverse algorithms to annotate the structural and functional attributes of the targeted fungus's HP [12].

In order to define the unknown proteome conformation, a number of computational approaches were utilized, including homology modeling, multiple sequence alignment, secondary, tertiary structure, or/and quality assessment. Additionally, important investigations such as phylogeny analysis, active site identification, and physiochemical characteristics was provided functional annotation of HP that helps to understand its role in metabolism, and also helps to identify new therapeutic targets [13, 14].

The goals of this present research involve utilizing bioinformatics tools to annotate the structural and functional features of an HP (accession number QWW2297.1) from *C. auris*. that aid in understanding the as-yet-unknown functions of this HP. The knowledge gained from this research will aid in the development of new drugs to combat this potential threat to humanity.

METHODS

Sequence Retrieval

The amino acid sequence of the HP (accession number QWW22972.1) originating from *C. auris* was obtained in FASTA format from the NCBI protein database [15]. Various standard prediction servers were utilized to annotate the obtained protein sequences, as detailed in **Table 1** [16].

Physicochemical Properties Analysis

The ExPASy ProtParam tool was employed to examine the physicochemical attributes of the HP. This examination encompassed parameters such as molecular weight, theoretical isoelectric point (pI), amino acid composition, count of positive and negative residues, instability index (II), aliphatic index (AI), grand average of hydropathicity (GRAVY), molecular formula, and estimated half-life [17].

Subcellular Localization Prediction

Sub-cellular localization of a protein is an essential for determining for the vaccine candidate [18]. Several tools were used for localizing the HP such as PSORT II [19], WolF PSORT [20], and Euk-mPLoc 2.0 [21].

Functional Annotation

Proteins are complex molecule that perform various functions in living organisms. They are categorized into distinct families and superfamilies according to their sequence attributes, domains, motifs, and functional resemblances [22]. The target protein's potential function and conserved domains were determined using NCBI conserved domain search service (CD search) [23], ScanProsite tool [24], InterPro [25], and SMART tool [26]. The folding pattern of the protein was detected through the utilization of the PFP-FunD SeqE method [27].

Multiple Sequences Alignment and Phylogenetic Tree Analysis

In order to find the protein's homologs, a BLASTp search was executed on the NCBI website against the nonredundant database, utilizing the default settings [29]. Following this, a multiple sequence alignment and a phylogenetic tree were produced using CLC Sequence Viewer version 8. The CLC Sequence Viewer software is commonly used for sequence analysis and visualization, aiding in the comparative analysis of protein sequences and the construction of the phylogenetic relationship.

Secondary Structure Prediction

The anticipated two-dimensional arrangement of the HP was forecasted through the application of the self-optimized prediction method with alignment (SOPMA) [28]. Additionally, the findings from SOPMA were corroborated using another tool called PSIPRED [30].

Homology Modeling

The SWISS-MODEL server was employed to ascertain the three-dimensional (3D) arrangement of the target protein via homology modeling [31]. This server independently conducts a BLASTp search to identify probable templates for each protein sequence. From the search results, the template protein 2×19.1.A was selected for homology modeling. This particular template, featuring an X-ray diffraction model of a GTP-binding nuclear protein, exhibited a substantial sequence identity of 94.15%, which proved to be a reliable basis for the modeling process. The visualization of the 3D model structure was carried out using PyMOL v2.0.

Model Quality Assessment

The PROCHECK [32], Verify3D [33], ERRAT [34], and modules from the SAVES server were employed to evaluate the validity of the projected 3D structure. Furthermore, the ExPASy server's QMEAN programs [35] within the SWISS-MODEL Workspace were utilized to compute the QMEAN Z-score and appraise the model's integrity. Additionally, the ProSA-web server [36] was employed to compute Z scores for both the template and target proteins.

Table 2. Physiochemical properties of HP calculated by ProtParam tool

Description	Value
Number of amino acids	215
Molecular weight (Da)	24,459.00
Theoretical pI	6.53
Number of positively charged residues	27
Number of negatively charged residues	28
Number of atoms	3443
Instability index	27.14
Aliphatic index	82.47
Grand average of hydropathicity (GRAVY)	-0.392

Protein-Protein Interaction Analysis

STRING is a useful database that offers details on observed and anticipated protein interactions. The database combines data from several sources to generate vast protein interaction networks. The genetic background, (conserved) co-expression patterns, high throughput experimental data, and previous understanding of protein interactions were a few examples of these sources [37].

Active Site Determination

For pinpointing the protein's active site, the computed atlas of surface topography of proteins (CASTp) server was harnessed [38]. CASTp furnishes detailed, comprehensive, and quantifiable insights into a protein's topographical attributes. It facilitates the accurate detection and measurement of active pockets located on protein surfaces as well as within the interior of 3D structures. Consequently, it has transformed into an indispensable tool for predicting the regions of a protein that engage with ligands. The outcome from CASTp was also visualized utilizing the PyMOL software.

Energy Minimization of the Model Structure

The 3D model structure derived from the SWISS-MODEL server underwent energy minimization with the YASARA [39] force field minimizer. This method reduces energy usage and imparts a more precise and stable 3D shape to the target protein.

RESULTS

Physiochemical Properties Analysis

Using the ProtParam tool, various physicochemical characteristics of the HP (accession no. QWW22972.1) were assessed and tabulated (**Table 2**). This protein comprised 215 amino acids and had a molecular weight of 24459.00 Da. Among its constituent amino acids, the most abundant were Ala (13), Arg (9), Asn (10), Asp (14), Cys (3), Gln (10), Glu (14), Gly (16), His (5), Ile (9), Leu (19), Lys (18), Met (4), Phe

Table 3. The BLASTp outcome indicating the degree of similarity among proteins

Accession No	Organisms	Protein name	TS	PI	E-value
QWW22972.1	*Candida auris*	Hypothetical protein			
XP_024714986.1	*Candida pseudohaemulonii*	GTP-binding nuclear protein GSP1/Ran	450	100%	1e-159
XP_002618167.1	*Clavispora lusitaniae* ATCC 42720	GTP-binding nuclear protein GSP1/Ran	449	100%	2e-159
XP_001527056.1	*Lodderomyces elongisporus* NRRL YB-4239	GTP-binding nuclear protein GSP1/Ran	444	100%	4e157
KGU32723.1	*Candida albicans* P75063	GTP-binding nuclear protein GSP1/Ran	439	99%	2e-155
XP_018714336.1	*Metschnikowia bicuspidata var. bicuspidata* NRRL YB-4993	GTP-binding nuclear protein GSP1/Ran	437	96.28%	1e-154
XP_020065809.1	*Suhomyces tanzawaensis* NRRL Y-17324	GTP-binding nuclear protein GSP1/Ran	434	97.63%	2e-153

Note. TS: Total score & PI: Percent identity

(11), Pro (11), Ser (4), Thr (14), Trp (3), Tyr (9), and Val (19). Calculations revealed the presence of 28 negatively charged residues (Asp + Glu) and 27 positively charged residues (Arg + Lys). The protein's isoelectric point (pI) was determined to be 6.53. With an II of 27.14, the HP exhibited stability. The AI of 82.47 indicated its stability across a broad temperature range. The GRAVY score (-0.392) indicated the protein's hydrophilic nature, making it soluble in water. The HP had a chemical formula of C1107H1717N293O319S7. Its estimated half-life (HL) was calculated to be > 20 h in yeast (in vivo), > 10 h in *Escherichia coli* (in vivo), and 30 h in human reticulocytes (in vitro). The protein's capacity to absorb light at a specific wavelength was denoted by its extinction coefficient value (30,035). It is posited that deriving the molar extinction coefficient from amino acid composition is feasible.

Subcellular Localization Prediction

Automated forecasting of protein subcellular localization is a vital technique in protein examination and annotation, as it offers insights into the role of a protein within an organism. Identification of the location of the protein within a cell is known as subcellular localization. The location of proteins within cells has a significant impact on how those proteins function. This data can also be utilized for the development of a medication aimed at the specific protein. Our target protein's subcellular location was predicted by the PSORT II, WolF PSORT algorithm to be "cytoplasmic." The protein was similarly identified as a cytoplasmic one by the Euk-mPLoc 2.0 protein subcellular localization server.

Functional Annotation

The protein's domain was estimated using a number of annotation tools, including NCBI-CD Search, ScanProsite, InterPro, and SMART. All the methods were indicated the presence of a domain from the nuclear superfamily of GTP-binding proteins known as the Ran family, which is a GTPase involved in importing and exporting proteins and RNAs from nucleus. The Ran family GTP-binding nuclear superfamily area was found in the protein sequence from amino acid positions 4 to 212, according to an investigation by the NCBI-CD search server. The E-value for this detection was 1.08e-162, indicate that there was a very strong match to this domain. The Ran family GTP-binding nuclear superfamily domain was found in the protein sequence, more precisely in the area starting at amino acid position 2 and ending at amino acid position 166, as predicted by ScanProsite. A domain in the protein with an E-value of 8.98e-154 that resembles the Ran family GTP-binding nuclear superfamily was identified by SMART analysis in a similar manner. This domain was located at position 14-212. These investigations demonstrated that the protein has a domain that belongs to the Ran family GTP-binding nuclear superfamily. The homogeneity across various prediction techniques further supports the existence of this domain in the protein sequence. The utilization of the PFP-FunDSeqE tool for recognizing fold patterns exposed the existence of a "P-loop" fold within the protein sequence. The P-loop-fold domain exhibits a three-layer αβα sandwich structure, with the P-loop being notably recognized for its occurrence in ATP- and GTP-binding proteins. Moreover, it's also identified in a range of proteins associated with phosphorylated substrates.

Multiple Sequence Alignment and Phylogenetic Tree Analysis

A BLASTp search conducted on the non-redundant database unveiled homology, with sequence similarities of up to 96%, between the protein in question and several established GTP-binding nuclear proteins. This includes well-known proteins like GSP1/Ran from various organisms (**Table 3**).

To discern the shared and distinct residues among these homologs, multiple sequence alignments were carried out for a selected subset of proteins based on the outcomes of the BLASTp analysis (**Figure 1**).

Using the identical data, a phylogenetic tree was formed (**Figure 2**). The target protein appears to have a common ancestor with two other proteins from *Candida pseudohaemulonii* and *Clavispora lusitaniae* ATCC 42720. The line segment with the number (0.010) on it represents how much genetic change has occurred as estimated by the scale bar.

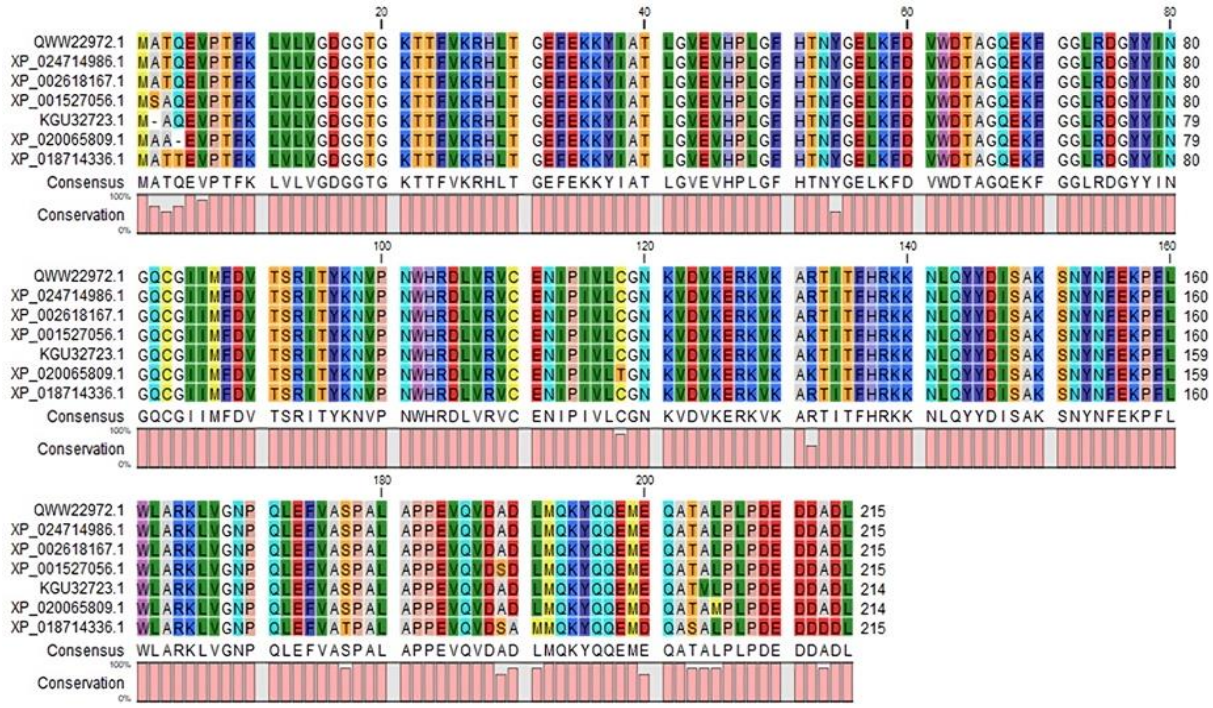

Figure 1. Multiple sequence alignments among different GTP-binding nuclear proteins using CLC Sequence Viewer version 8 (Top row-target protein, row 2-*Candida pseudohaemulonii*, row 3-*Clavispora lusitaniae* ATCC 42720, row 4-*Lodderomyces elongisporus* NRRL YB-4239, row 5-*Candida albicans* P75063, row 6-*Suhomyces tanzawaensis* NRRL Y-17324, and row 7-*Metschnikowia bicuspidata var. bicuspidata* NRRL YB-4993) (Source: Authors' own elaboration; organism names are obtained from Blast Result; and MSA obtained by CLC sequence viewer version 8)

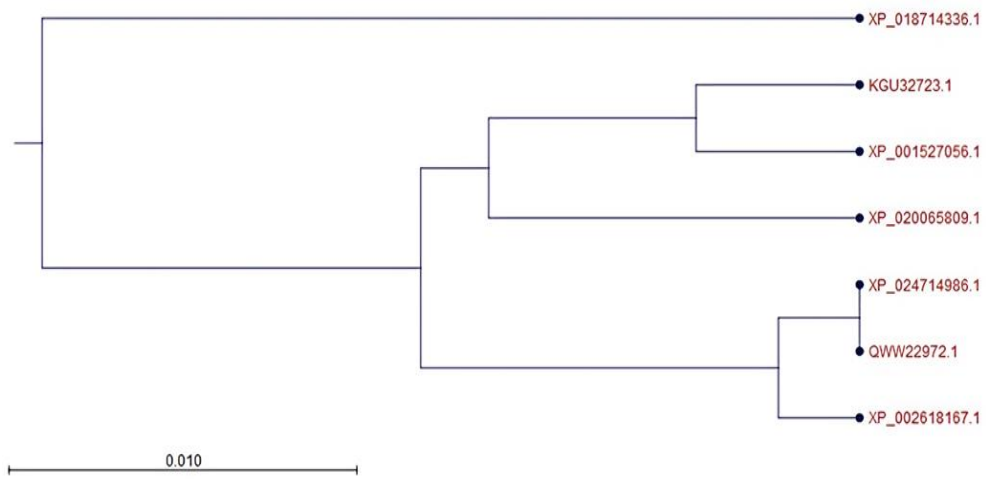

Figure 1. A phylogenetic tree was constructed to depict evolutionary connection between target protein & other GTP-binding nuclear proteins (This tree was generated using CLC Sequence Viewer version 8, scale bar provides an estimation of sequence divergence, & extent of genetic variation is symbolized by the line segment along with its corresponding value [0.010]) (Source: Authors' own elaboration)

Secondary Structure Prediction

The structure of a protein largely preserves its function. A notable segment of the protein's secondary structure consists of helices, sheets, turns, and coils. The secondary structure of the HP, as assessed by the SOPMA server, displayed an alpha helix content of 55.16%, random coil of 33.41%, extended strand of 7.17%, and beta-turn of 4.26% (**Figure 3**). This outcome was corroborated by a similar result obtained from the PSIPRED server (**Figure 4**).

Tertiary Structure (3D) Prediction and Model Quality Assessment

The tertiary structure of the target protein was acquired through the utilization of the template 2×19.1.A, sourced from the SWISS-MODEL platform, showcasing a high sequence similarity (94.15%) with the target protein. PyMOL 2.0 was employed to visually depict the protein's 3D arrangement (**Figure 5**).

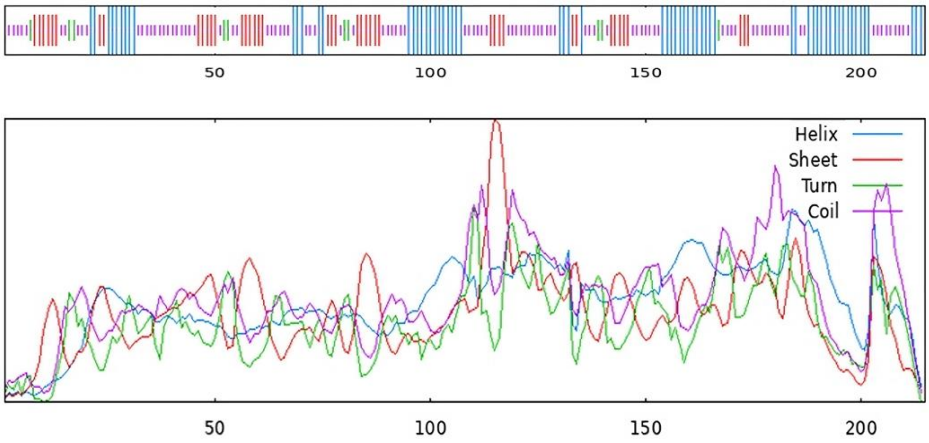

Figure 3. Secondary structure predicted by the SOPMA server (Source: Authors' own elaboration)

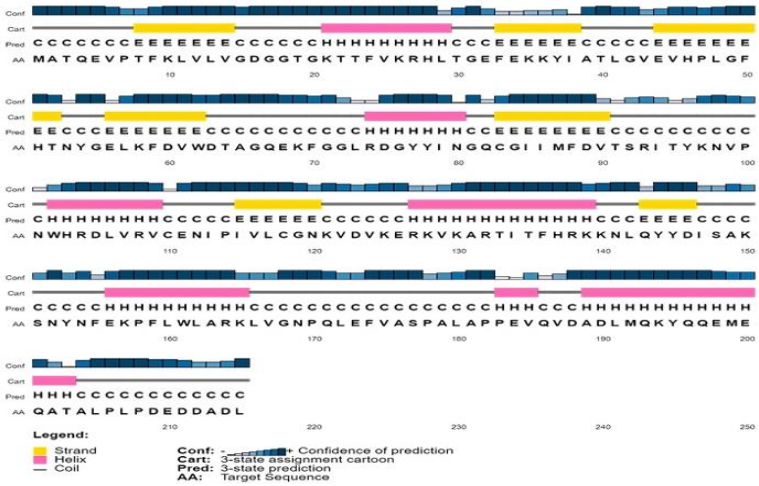

Figure 4. Anticipated secondary structure of target protein was determined through utilization of PSI-PRED server (There are four distinct sections in this graphical illustration: the first section consists of bars with different heights, the confidence score is proportional to the bar height's length, the alpha helix, beta sheets or strands, and coils are represented in the second section by the colors pink, yellow, and gray, respectively, a coil links a certain alpha helix to a specific set of beta sheets, the secondary structure of a protein is shown in the third part using an alphabetic form, here, beta sheets, alpha helixes, and coils are represented by letters E, H, and C, respectively, & organization of amino acids is shown in alphabetical order in the last section) (Source: Authors' own elaboration)

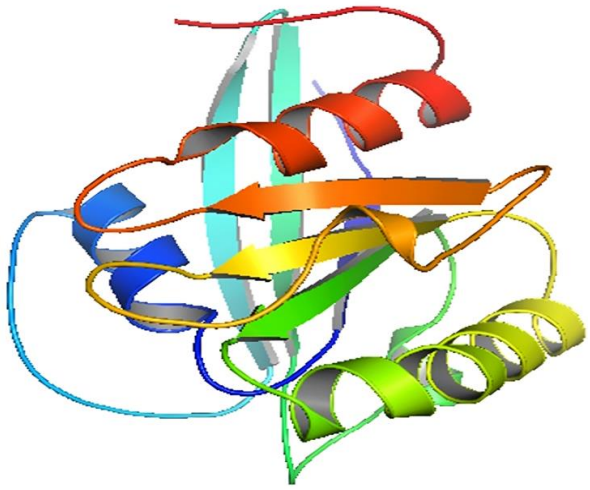

Figure 5. The projected three-dimensional structure of the target protein was obtained using the SWISS-MODEL server and then visualized using PyMOL software (Source: Authors' own elaboration)

The fidelity of the 3D model was assessed employing various evaluation tools, including PROCHECK, Verify3D, ERRAT, and QMEAN. The PROCHECK results demonstrated that a substantial 90.5% of amino acid residues resided within the most favored region of the Ramachandran plot, substantiating the model's sound quality (**Table 4** and part A in **Figure 6**).

In terms of the 3D plot verification, it was determined that 86.55% of the residues exhibited an average 3D-1D score ≥ 0.2. The ERRAT assessment confirmed the model's high quality, boasting an overall quality factor of 94.410 (part B in **Figure 6**). The QMEAN4 score of -0.51 was attributed to the model, placing it in the dark grey region of the QMEAN analysis (part C in **Figure 6**), indicating an acceptable level of quality. The model's Z score was employed to assess its overall quality, checking if the input structure fell within the score range typical for native proteins of comparable size. Both the template and model Z-scores from the ProSA

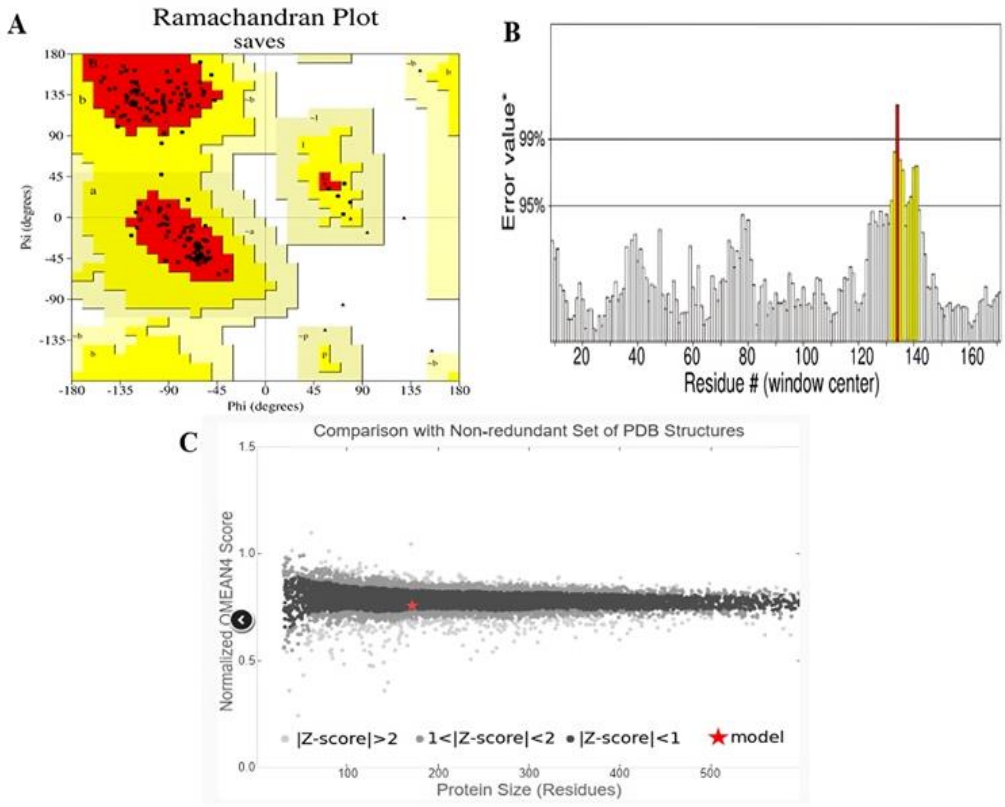

Figure 6. Assessment of model quality: (A) validation of the model structure was performed using the PROCHECK server, and the resulting Ramachandran plot indicated that 90.5% of amino acid residues were situated within the most favored regions [A, B, L], (B) the ERRAT output yielded a quality factor of 94.410, the two lines on the error axis indicate the threshold for rejecting areas that surpass the error value, & (C) a graphical representation of the QMEAN analysis for the model structure revealed a Z score of -0.51 (this Z score signifies a strong concordance between the model's structure and experimental structures of similar size) (Source: Authors' own elaboration)

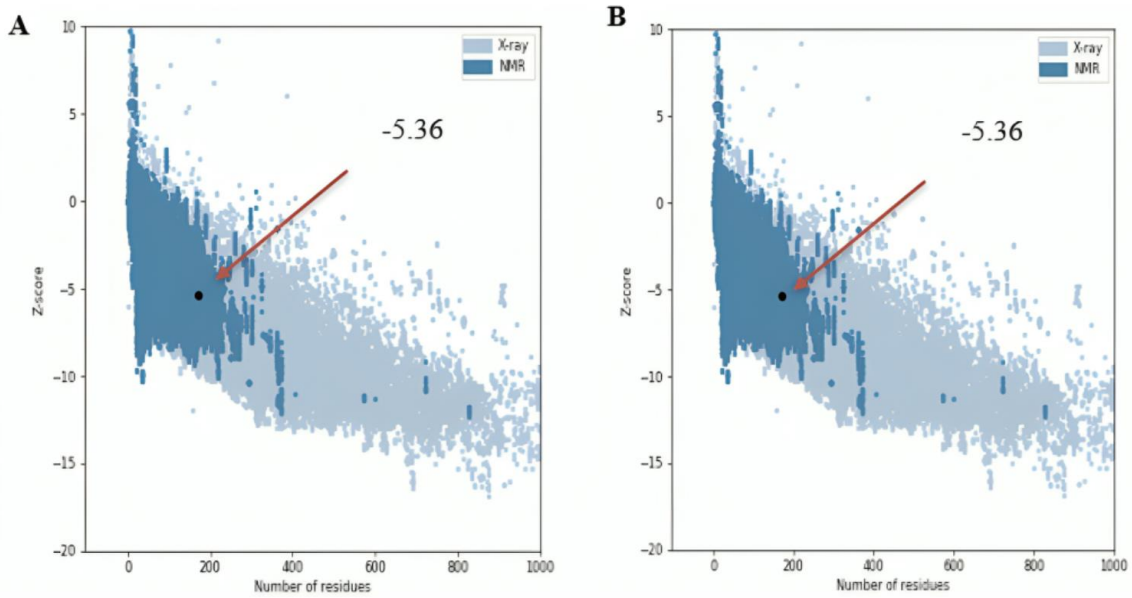

Figure 7. Z-scores were computed for both target protein (A) and template protein (B) using ProSA server (notably, both structures were situated within the range commonly observed for native proteins of similar size, including those determined through experimental techniques such as NMR & X-ray) (Source: Authors' own elaboration)

analysis were recorded as -5.36 (part A and part B in **Figure 7**), signifying homology between the template and the model.

Protein-Protein Interaction Analysis

The STRING tool was predicted the interaction of selected HP with 10 proteins. A0A2H1A6Z0 was the most

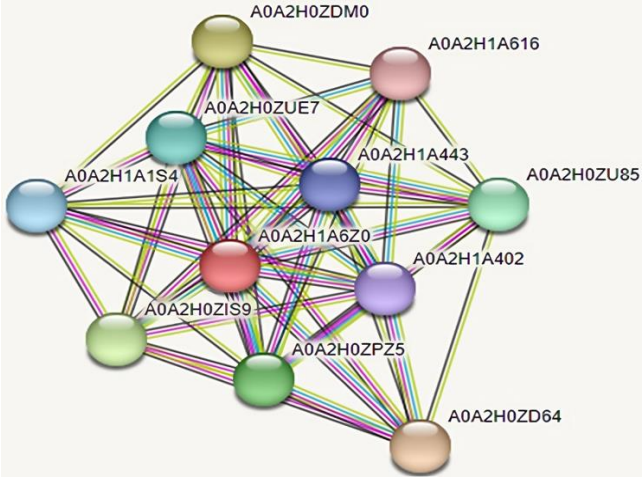

Figure 8. The interaction network between proteins for the hypothetical protein was generated using the STRING server (Source: Authors' own elaboration)

query protein. The recognized functional associates accompanied by their corresponding scores were A0A2H0ZD64 (0.999) which encodes a ubiquitin-like protein SMT3, A0A2H0ZDM0 (0.999) that encodes an exportin-T family protein, A0A2H0ZIS9 (0.999) that encodes an elongation factor 1-alpha protein, A0A2H0ZPZ5 (0.999), A0A2H0ZU85 (0.999), A0A2H1A443 (0.999), A0A2H1A402 (0.996) that encodes an uncharacterized protein, A0A2H0ZUE7 (0.999) that encodes a nuclear transport factor 2 protein, A0A2H1A1S4 (0.999) that encodes Xpo1 domain-containing protein, and A0A2H1A616 (0.996) that encodes MIF4G domain-containing protein (**Figure 8**).

Active Site Determination

The active site analysis of the developed 3D model structure was performed using the CASTp server. A depiction of the amino acid residues comprising the active site is presented in part A and part B in **Figure 9**.

Analysis of the protein's predicted active site revealed the participation of 15 distinct amino acids forming a potent active site region. One of the biggest pockets, with a total volume of 128.648 amino acids and 155.344 solvent accessible (SA) surface area that contained the most active site. ASP[16], THR[64], ALA[65], GLY[66], GLN[67], GLU[68], GLY[71], ARG[74], ASP[75], TYR[77], TYR[78], ASN[98], ASN[101], TRP[102], and ASP[105] were the important active residues that were predicted from the pocket. The recognition and detailed characterization of these active site residues constitute a pivotal stage in the process of designing drugs or inhibitors.

Energy Minimization Result

The YASARA force field minimizer was employed to optimize the energy of the protein's 3D structure. Using the YASARA energy minimization server, the energy of the model protein structure was reduced from -80,417.4 kJ/mol to -108,599.6 kJ/mol (as shown in **Figure 10**). After energy reduction, the final score was changed to -0.54, suggesting a more stable shape, from the original value of -1.20.

DISCUSSION

C. auris is an emerging yeast that exhibits antibiotic resistance and is associated with invasive infections, presenting a substantial global health issue due to its high mortality rate [40]. Till to date, there is no vaccine for *C. auris* [41]. Though research on HPs has not yet kept up with the rapid development of low-cost sequencing technology, there is now an enormous quantity of genomic and proteomic data available [42]. HPs make up a large percentage of genomes and frequently lack thorough functional characterization or annotation. Researchers or scientists can gain a deeper understanding of cellular operations by carefully annotating HPs to reveal their significance in diverse biological processes and pathways [43]. The characterization of HPs helps to explore the

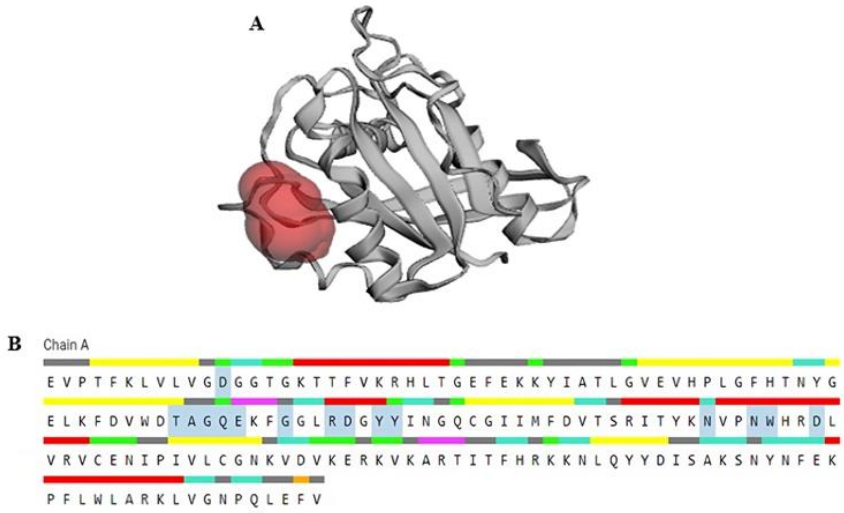

Figure 9. Identification of active site using CASTp server (A) (the largest active site was found in the areas with 155.344 and a volume of 128.648 amino acids) & active amino acid residues are highlighted in (B) (Source: Authors' own elaboration)

Figure 10. Energy minimization of model (YASARA energy minimized server) & result is visualized (YASARA View software) (Source: Authors' own elaboration)

knowledge of organisms' metabolic routes, the understanding of disease advancement, the advancement of pharmaceuticals, and the formulation of strategies for disease management [42].

In this research, several highly effective bioinformatic tools and databases were employed to characterize the HP (accession no. QWW22972.1) of *C. auris* from structural and functional perspectives. With a GRAVY score of -0.392 and an II of 27.14, the protein was determined to have 215 amino acids based on physiochemical data, indicating that it is polar and stable (**Table 2**). The predictions of PSORT II, WolF PSORT, and Euk-mPLoc 2.0 servers indicated that the protein was localized in the cytoplasm.

The CD Search and ScanProsite, InterPro, and SMART studies were unveiled that the HP belonged to the nuclear superfamily of Ran family GTP-binding proteins, according to domain and motif analysis. Ran, a small nuclear GTPase, strictly regulates the movement of large molecules between the nuclei and cytosol. Due to the substantial remodelling, the nucleus goes through for proper chromosomal segregation during mitosis, the protein is particularly important for sustaining the nucleus [44]. Additionally, Ran plays a crucial role in tasks such as the construction of the mitotic spindle, regulation of nuclear envelope dynamics, and precise coordination of cell cycle transitions [45-47]. However, these processes shed light on the temporal and spatial changes in cellular structure that take place during the cell division cycle, particularly during interphase [48]. Ran GTP, which was served as a chemical signal, also points to the location of the nucleus. Importins and exportins were transport molecules that operate as Ran's effectors. They facilitate the movement of cargo between the nucleus and cytoplasm [49]. As a result, Ran GTP accurately and effectively regulates how importins and exportins interact with their respective cargoes, ensuring that the cargoes are transported in the desired direction [50]. Performing a BLASTp search against the non-redundant database unveiled a sequence similarity of up to 96% with other GTP-binding nuclear proteins, notably including the GSP1/Ran protein from various organisms (**Table 3**). To see the conserved and different residues among the homologs, multiple sequence alignments of a chosen few proteins were performed using the BLASTp findings (**Figure 1**). Lastly, the gathered data was utilized to construct a phylogenetic tree (**Figure 2**).

According to the inspection of its secondary structure, the protein has a random coil, beta-turn, alpha helix, and extended strand. The alpha helix was the most predominant of these elements (**Figure 3** and **Figure 4**). In addition, the 3D structure of the protein depicted from SWISS-MODEL server successfully passed the tools used to evaluate quality, demonstrating that the overall level of quality was good. The model's amino acid environment was examined using the VERIFY 3D program, and it was discovered that 86.55% of the residues had an average 3D-1D score of 0.1, suggesting an ambient environment, for non-bonded atomic interactions. Additionally, the ERRAT total quality factor of 94.410 indicated a high-quality model (part B in **Figure 6**). Interestingly, the Ramachandran plot was represented most of the amino acid residues (90.5%) in its favored region and zero percentage of disallowed region that indicates the validity of the model's quality (**Table 4** and part A in **Figure 6**). It is generally acknowledged that a 3D model is credible if more than 90% of the residues are in the regions that are preferred [51]. The QMEAN4 server's output showed that the predicted model's Z-score was -0.51, which likewise indicates a high-quality model (part C in **Figure 6**). The Z-score serves as an indicator of the overall quality of the model and is utilized to verify whether the input structure falls within the range of scores typically associated with native proteins of similar size [36]. Additionally, the Z-score for both the model and the template, as obtained from ProSA, was -5.36 (part A and part B in **Figure 7**), suggesting a homologous relationship between the template and the model.

At the same time, the protein-protein interaction network explored some neighboring proteins that was mostly uncharacterized (**Figure 8**). The proteins having a score of 0.996 and 0.999 meaning that all the proteins were close to the target protein. Most importantly, ubiquitin-like protein SMT3, elongation factor 1-alpha protein, and nuclear transport factor 2 protein was the predominant but only MIF4G domain-containing protein was regulated in

Table 4. Ramachandran plot statistics of the target protein

Statistics	Number of AA residues	Percentage (%)
Residues in the most favored region [A, B, L]	133	90.5%
Residues in the additional allowed regions [a, b,l,p]	14	9.5%
Residues in the generously allowed regions [~ a, ~b, ~l,~p]	0	0.0%
Residues in disallowed regions	0	0.0% (Total = 100%)
Number of non-glycine and non-proline residues	147	
Number of end-residues (excl. Gly & Pro)	3	
Number of glycine residues	16	
Number of Proline residues	6	
Total number of residues	172	

p27-dependent cell proliferation in host cell at least functional activity [52].

The location of the active site on a protein holds significant importance for several purposes, such as identifying structural features, comparing functional sites, conducting molecular docking, and even designing new drugs [53]. During the CASTp analysis, the projected active site of the protein revealed the involvement of 15 amino acids in a potent active site. Within this context, one notably large pocket was identified as an active site, featuring an SA surface area of 155.344 and a volume encompassing 128.648 amino acids (part A and part B in **Figure 9**). Following the YASARA energy minimization process, the 3D structure of the target protein achieved increased stability, leading to a score of -0.54 (**Figure 10**). Importantly, these investigations into the annotation of this HP have facilitated the development of potent drugs and vaccines against pathogens of this nature.

CONCLUSION

Several bioinformatics methods were employed in this work to analyze a HP from the fungus *C. auris*. Our research has implications for enhancing the target protein's functioning and raising resource usage effectiveness. Future research aims to experimentally test our results and generate novel ligands for medication development using structural and functional data. Future treatment plans must be improved by continued study of target proteins and their effectors in *C. auris* and other species. We may learn more from this work about investigating the structural and functional studies of proteins with unidentified activities. In the future, other researchers may use the results of this study to conduct their own *in-silico* research.

Author contributions: NA, THE, MAH, MAM, & MEM: acquisition, investigation, curation, analysis, or interpretation of data; **NA & THE:** methodology; **NA:** writing the first draft; **MAH, FA, MAK, MNI, & MEM:** writing–review & editing; & **MEM:** conceptualization or design. All authors have agreed with the results and conclusions.

Funding: No funding source is reported for this study.

Acknowledgments: The authors would like to thank the Department of Biotechnology and Genetic Engineering at Noakhali Science and Technology University, Noakhali-3814, Bangladesh, for providing opportunities to conduct this study.

Ethical statement: The authors stated that the study does not require any ethical approval since the study solely involves modeling, or analysis using pre-existing data, algorithms, or mathematical frameworks without any direct interaction with human or animal subjects, or the use of sensitive or personal data.

Declaration of interest: No conflict of interest is declared by the authors.

Data sharing statement: Data supporting the findings and conclusions are available upon request from the corresponding author.

REFERENCES

1. Du H, Bing J, Hu T, Ennis CL, Nobile CJ, Huang G. Candida auris: Epidemiology, biology, antifungal resistance, and virulence. PLoS Pathog. 2020;16(10): e1008921. doi:10.1371/journal.ppat.1008921
2. Ahmad S, Alfouzan W. Candida auris: Epidemiology, diagnosis, pathogenesis, antifungal susceptibility, and infection control measures to combat the spread of infections in healthcare facilities. Microorganisms. 2021;9(4):807. doi:10.3390/microorganisms9040807
3. Spivak ES, Hanson KE. Candida auris: An emerging fungal pathogen. J Clin Microbiol. 2018;56(2):e01588-17. doi:10.1128/JCM.01588-17
4. Chowdhary A, Sharma C, Meis JF. Candida auris: A rapidly emerging cause of hospital-acquired multidrug-resistant fungal infections globally. PLoS Pathog. 2017;13(5):e1006290. doi:10.1371/journal.ppat.1006290
5. Khan Z, Ahmad S. Candida auris: An emerging multidrug-resistant pathogen of global significance. Curr Med Res Pract. 2017;7(6):240-8. doi:10.1016/j.cmrp.2017.11.004
6. Satoh K, Makimura K, Hasumi Y, Nishiyama Y, Uchida K, Yamaguchi H. Candida auris sp. nov., a novel ascomycetous yeast isolated from the external ear canal of an inpatient in a Japanese hospital. Microbiol Immunol. 2009;53(1):41-4. doi:10.1111/j.1348-0421.2008.00083.x

7. Vallabhaneni S, Kallen A, Tsay S, et al. Investigation of the first seven reported cases of Candida auris, a globally emerging invasive, multidrug-resistant fungus-United States, May 2013-August 2016. Morb Mortal Wkly Rep. 2016;65(44):1234-7. doi:10.15585/mmwr.mm6544e1
8. Sanyaolu A, Okorie C, Marinkovic A, et al. Candida auris: An overview of the emerging drug-resistant fungal infection. Infect Chemother. 2022;54(2):236. doi:10.3947/ic.2022.0008
9. Ademe M, Girma F. Candida auris: From multidrug resistance to pan-resistant strains. Infect Drug Resist. 2020;13:1287-94. doi:10.2147/IDR.S249864
10. Yadav A, Singh A, Wang Y, et al. Colonisation and transmission dynamics of Candida auris among chronic respiratory diseases patients hospitalised in a chest hospital, Delhi, India: A comparative analysis of whole genome sequencing and microsatellite typing. J Fungi. 2021;7(2):81. doi:10.3390/jof7020081
11. Shahbaaz M, Imtaiyaz Hassan M, Ahmad F. Functional annotation of conserved hypothetical proteins from Haemophilus influenzae Rd KW20. PloS One. 2013;8(12):e84263. doi:10.1371/journal.pone.0084263
12. Ijaq J, Chandra D, Ray MK, Jagannadham MV. Investigating the functional role of hypothetical proteins from an antarctic bacterium Pseudomonas sp. Lz4W: Emphasis on identifying proteins involved in cold adaptation. Front Genet. 2022;13:825269. doi:10.3389/fgene.2022.825269
13. 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(1-2):90-127. doi:10.1016/j.pneurobio.2005.10.001
14. Galperin MY, Koonin EV. 'Conserved hypothetical' proteins: Prioritization of targets for experimental study. Nucleic Acids Res. 2004;32(18):5452-63. doi:10.1093/nar/gkh885
15. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL. GenBank. Nucleic Acids Res. 2003;31:23-7. doi:10.1093/nar/gkg057
16. Rabbi MF, Akter SA, Hasan MJ, Amin A. In silico characterization of a hypothetical protein from *Shigella dysenteriae* ATCC 12039 reveals a pathogenesis-related protein of the type-VI secretion system. Bioinform Biol Insights. 2021;15:11779322211011140. doi:10.1177/11779322211011140
17. Wilkins MR, Gasteiger E, Bairoch A, et al. Protein identification and analysis tools on the ExPASy server. Methods Mol Biol. 1999;112:531-52. doi:10.1385/1-59259-584-7:531
18. Vetrivel U, Subramanian G, Dorairaj S. A novel in silico approach to identify potential therapeutic targets in human bacterial pathogens. Hugo J. 2011;5(1-4):25-34. doi:10.1007/s11568-011-9152-7
19. Nakai K, Horton P. PSORT: A program for detecting sorting signals in proteins and predicting their subcellular localization. Trends Biochem Sci. 1999;24(1):34-5. doi:10.1016/S0968-0004(98)01336-X
20. Horton P, Park KJ, Obayashi T, et al. WoLF PSORT: Protein localization predictor. Nucleic Acids Res. 2007;35(Web Server issue): W585-7. doi:10.1093/nar/gkm259
21. Chou K-C, Shen H-B. A new method for predicting the subcellular localization of eukaryotic proteins with both single and multiple sites: Euk-mPLoc 2.0. PLoS One. 2010;5(4):e9931. doi:10.1371/journal.pone.0009931
22. Wu CH, Huang H, Yeh LS, Barker WC. Protein family classification and functional annotation. Comput Biol Chem. 2003;27(1):37-47. doi:10.1016/S1476-9271(02)00098-1
23. Lu S, Wang J, Chitsaz F, et al. CDD/SPARCLE: The conserved domain database in 2020. Nucleic Acids Res. 2020;48(D1):D265-8. doi:10.1093/nar/gkz991
24. de Castro E, Sigrist CJ, Gattiker A, et al. ScanProsite: Detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. Nucleic Acids Res. 2006;34(Web Server issue):W362-5. doi:10.1093/nar/gkl124
25. Apweiler R, Attwood TK, Bairoch A, et al. The InterPro database, an integrated documentation resource for protein families, domains and functional sites. Nucleic Acids Res. 2001;29(1):37-40. doi:10.1093/nar/29.1.37
26. Letunic I, Doerks T, Bork P. SMART 7: Recent updates to the protein domain annotation resource. Nucleic Acids Res. 2012;40(D1):D302-5. doi:10.1093/nar/gkr931
27. Shen H-B, Chou K-C. Predicting protein fold pattern with functional domain and sequential evolution information. J Theor Biol. 2009;256(3):441-6. doi:10.1016/j.jtbi.2008.10.007
28. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403-10. doi:10.1016/S0022-2836(05)80360-2
29. Combet C, Blanchet C, Geourjon C, Deleage G. NPS@: Network protein sequence analysis. Trends Biochem Sci. 2000;25(3):147-50. doi:10.1016/S0968-0004(99)01540-6
30. Buchan DW, Jones DT. The PSIPRED protein analysis workbench: 20 years on. Nucleic Acids Res. 2019;47(W1):W402-7. doi:10.1093/nar/gkz297
31. Waterhouse A, Bertoni M, Bienert S, et al. SWISS-MODEL: Homology modeling of protein structures and complexes. Nucleic Acids Res. 2018;46(W1):W296-303. doi:10.1093/nar/gky427
32. Laskowski RA, Rullmann 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;8:477-86. doi:10.1007/BF00228148

33. Eisenberg D, Lüthy R, Bowie JU. VERIFY3D: Assessment of protein models with three-dimensional profiles. In: Methods in enzymology. Academic Press; 1997. p. 396-404. doi:10.1016/S0076-6879(97)77022-8
34. Colovos C, Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci. 1993;2(9):1511-9. doi:10.1002/pro.5560020916
35. Benkert P, Biasini M, Schwede T. Toward the estimation of the absolute quality of individual protein structure models. Bioinform. 2011;27(3):343-50. doi:10.1093/bioinformatics/btq662
36. Wiederstein M, Sippl MJ. ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res. 2007;35(Web Server issue):W407-10. doi:10.1093/nar/gkm290
37. 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. 2010;39(Database issue):D561-8. doi:10.1093/nar/gkq973
38. Tian W, Chen C, Lei X, Zhao J, Liang J. CASTp 3.0: Computed atlas of surface topography of proteins. Nucleic Acids Res. 2018;46(W1):W363-7. doi:10.1093/nar/gky473
39. Krieger E, Joo K, Lee J, et al. Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. Proteins: Struct Funct Bioinform. 2009;77(S9):114-22. doi:10.1002/prot.22570
40. Morales-López SE, Parra-Giraldo CM, Ceballos-Garzón A, et al. Invasive infections with multidrug-resistant yeast Candida auris, Colombia. Emerg Infect Dis. 2017;23(1):162. doi:10.3201/eid2301.161497
41. Pandey J. Clinical significance and vaccine development for Candida auris. Virol Mycol. 2022;11:240.
42. Sen T, Verma NK. Functional annotation and curation of hypothetical proteins present in a newly emerged serotype 1c of Shigella flexneri: Emphasis on selecting targets for virulence and vaccine design studies. Genes (Basel). 2020;11(3):340. doi:10.3390/genes11030340
43. Kumar K, Prakash A, Tasleem M, Islam A, Ahmad F, Hassan MI. Functional annotation of putative hypothetical proteins from Candida dubliniensis. Gene. 2014;543(1):93-100. doi:10.1016/j.gene.2014.03.060
44. Clarke PR, Zhang C. Spatial and temporal coordination of mitosis by Ran GTPase. Nat Rev Mol Cell Biol. 2008;9(6):464-77. doi:10.1038/nrm2410
45. Carazo-Salas RE, Gruss OJ, Mattaj IW, Karsenti E. Ran-GTP coordinates regulation of microtubule nucleation and dynamics during mitotic-spindle assembly. Nat Cell Biol. 2001;3(3):228-34. doi:10.1038/35060009
46. Mühlhäusser P, Kutay U. An in vitro nuclear disassembly system reveals a role for the RanGTPase system and microtubule-dependent steps in nuclear envelope breakdown. The J Cell Biol. 2007;178(4):595-610. doi:10.1083/jcb.200703002
47. Moore JD. The Ran-GTPase and cell-cycle control. Bioessays. 2001;23(1):77-85. doi:10.1002/1521-1878(200101)23:1<77::AID-BIES1010>3.0.CO;2-E
48. Mattila PK, Lappalainen P. Filopodia: Molecular architecture and cellular functions. Nat Rev Mol Cell Biol. 2008;9(6):446-54. doi:10.1038/nrm2406
49. Hetzer M, Gruss OJ, Mattaj IW. The Ran GTPase as a marker of chromosome position in spindle formation and nuclear envelope assembly. Nat Cell Biol. 2002;4(7):E177-84. doi:10.1038/ncb0702-e177
50. Cavazza T, Vernos I. The RanGTP pathway: From nucleo-cytoplasmic transport to spindle assembly and beyond. Front Cell Dev Biol. 2016;3:82. doi:10.3389/fcell.2015.00082
51. Hooda V, Babu Gundala P, Chinthala P. Sequence analysis and homology modeling of peroxidase from Medicago sativa. Bioinformation. 2012;8(20):974-9. doi:10.6026/97320630008974
52. Rajasekaran S, Soundararajan DC, Nayagam SM, et al. Modic changes are associated with activation of intense inflammatory and host defense response pathways-molecular insights from proteomic analysis of human intervertebral discs. The Spine J. 2022;22(1):19-38. doi:10.1016/j.spinee.2021.07.003
53. Mazumder L, Hasan MR, Fatema K, Islam MZ, Tamanna SK. Structural and functional annotation and molecular docking analysis of a hypothetical protein from Neisseria gonorrhoeae: An in-silico approach. Biomed Res Int. 2022;2022:4302625. doi:10.1155/2022/4302625

