LGU Journal of
**LIFE SCIENCES**

*LGU Society of Life Sciences*

**DOI: https://doi.org/10.54692/lgujls.2024.0803xxx**
Paper Submission: 24th July 2024; Paper Acceptance: 7th Sep 2024; Paper Publication: 10th Sep 2024

Research Article
Vol 8 Issue 3 July- Sep 2024

LGU J. Life. Sci
ISSN 2519-9404
eISSN 2521-0130

# Structural and Functional Annotations of Hypothetical Proteins of *Candida auris* for Novel Drug Target Identification: An in-silico Approach

Qamar Abbas[1], Sana Batool[2], Muhammad Zaid[1], Mahnoor Mushtaq[1], Maham Ijaz[2], Ayman Naeem[1], Ali Munir[2], Arslan Hamid[3], Naeem Mahmood Ashraf[4], Hina Batool[1]*

1. Department of Life Sciences, University of Management and Technology, Pakistan
2. School of Biological Sciences, University of the Punjab, Lahore, Pakistan.
3. University of Bonn, Germany
4. School of Biochemistry and Biotechnology, University of the Punjab, Lahore, Pakistan

**Corresponding Author's Email: hinabatool198@gmail.com**

***ABSTRACT:*** *Candida auris is an emerging fungal pathogen that causes severe invasive infections in healthcare facilities which are difficult to control and treat due to its resistance to major antifungal drugs. A large fraction of C. auris proteins are uncharacterized and hypothetical, and structural and functional characterization of these proteins can aid in the selection of novel drug targets. The present study involves a computational approach for the structural and functional characterization of hypothetical proteins from Candida auris. After the sequence retrieval, hypothetical proteins were predicted for physiochemical properties and subcellular localization, structurally modeled using I-TASSER, quality assessed through Verify3D, ERRAT, and PROCHECK, and functionally annotated to reveal conserved domains and roles in pathogen pathways. Finally, the immunogenic assessments, non-human homologous analysis, and druggability analysis reveal three hypothetical proteins of C. auris (KND95408.2, KND95415.2, and KND95429.2) as novel drug targets. Furthermore, the stable conformations of these selected drug targets with the minimum root mean square deviations (RMSD) and fluctuation (RMSF) were analyzed by MD simulations. Conclusively, the functional annotations of these hypothetical proteins can help in understanding the disease mechanisms at the molecular level, as well as provide new targets for drug development against Candida auris.*

***Keywords:*** Candida auris, Drug targets, In silico analysis, Candidiasis

# INTRODUCTION

Recently, several fungal infections have posed significant health concerns (Almeida et al., 2019). *Candida auris* is an emerging fungus that presents a serious global health threat causing severe illness in hospitalized and immunocompromised patients (Du et al., 2020). According to recent statistics, *C. auris* infects thousands of individuals globally, and the infectivity rate has gradually increased over the last decade (Vila et al., 2020). The pathogen is associated with many invasive infections of the ear, urinary system, skin, bone, wounds, bloodstream, and lower respiratory tract (Horton and Nett, 2020). Since its first isolation in 2009 from the ear canal of an infected female in Japan, the infectious outbreaks of *C. auris* have covered 45 countries with a mortality rate of 72% (Cândido et al., 2020; D'Ambra, 2019; Dahiya et al., 2020). Due to the nosocomial origin of these infections, the principal sources for the outspread of this infection include contaminated surfaces in hospitals and other healthcare settings (Bandara and Samaranayake, 2022). The pathogen colonizes the patient's skin, and the mucosal surfaces, contaminating the surroundings, leading to rapid transmission into the environment (Büyüktuna et al., 2019).

The excessive use of antibiotics and antifungal agents, a recent history of surgery, prolonged stay in the Intensive Care Unit (ICU), use of indwelling lines, central venous catheters, and feeding tubes are the main risk factors associated with *C. auris* infections (Thomas-RŘddel et al., 2022). *Candida auris* can also co-infect individuals infected with other pathogens, as in the case of COVID-19, where infected patients who were on ventilators acquired coinfection of *C. auris*, complicating the diagnosis, treatment, and prognosis, and increasing disease symptoms and mortality (Calvo et al., 2021).

Several antifungal agents, including echinocandins, fluconazole, and polyenes, have been proven effective against *C. auris*. However, the resistance of *C. auris* to these agents has increased the infection rates leading to increased mortalities and morbidities (Fang et al., 2021). Consequently, there is a need to

develop new treatment options to curtail the increasing infection rate of *C. auris* (Ré et al., 2021). In several emerging pathogens, the molecular functions of some proteins are unknown, and these uncharacterized proteins are known as hypothetical proteins (HPs). The functional annotation of these uncharacterized proteins helps in understanding their roles in different pathways and to identify novel drug targets for these pathogens (Omeershffudin and Kumar, 2019). Thus, in addition to the known proteins, the hypothetical proteins of emerging pathogens need to be analyzed for their potential drug target abilities (Chirgadze et al., 2022).

Recently, the uncharacterized proteins from many deadly pathogens have been suggested as ideal drug candidates for controlling their infections (Abbasi et al., 2022). Several computational tools are available for the structural and functional annotations of HPs, physicochemical characterization, protein network and pathway analysis, and molecular interaction exploration (Prabhu et al., 2020). Moreover, the computational screening of uncharacterized proteins may help in probing the potential drug targets, followed by measuring their affinities to various therapeutic agents. Therefore, screening the proteome of drug-resistant pathogens for discovering novel drug targets would be a new strategy to cater to the increased infectivity rates of these pathogens (Shamsinejad et al., 2022). The whole genome analysis of *C. auris* indicates that this pathogen contains over 5500 genes. Currently, 8357 proteins are reported, out of which many proteins are hypothetical (Jain et al., 2022).

The present study involves a computational strategy for the structural and functional characterization of the hypothetical proteins of *Candida auris* and subsequently screening for potential drug target proteins. The selected proteins should be essential for the survival of the pathogen and should not be homologous to the human proteome. The conformational stabilities of selected drug

targets were determined using molecular dynamics simulations. This study suggests novel drug targets for the pathogen *C. auris*, which may help in designing new and effective treatment options against this deadly pathogen. Further experimental confirmation through wet-lab analysis is needed.

## MATERIALS AND METHODS

**Selection of HPs and their Sequence Retrieval**

The hypothetical proteins of *Candida auris* and their sequence in the FASTA format were retrieved from the NCBI database (https://www.ncbi.nlm.nih.gov/) (Schoch et al., 2020). After the sequence retrieval, the sequence homology analysis of hypothetical proteins was performed using the BLASTp server (Wu et al., 2019). The proteins having homology with the structurally and functionally characterized proteins were selected for the downstream analysis.

**Physiochemical Characterization of Selected HPs**

The physicochemical parameters of selected HPs, including molecular weight, extinction coefficient, theoretical pI, instability index, aliphatic index, and Grand Average of Hydropathy (GRAVY), were analyzed via the online Expasy ProtParam (https://www.web.expasy.org/protparam) (Duvaud et al., 2021).

**Subcellular Localization**

The cellular location of proteins provides information about their function and helps predict the protein-protein interactions, which reveals these proteins' involvement in various signalling pathways (Wang and Chen, 2022). Based on this, the subcellular localization of selected hypothetical proteins was assessed using the CELLO2GO (http://cello.life.nctu.edu.tw/cello2go/) online server. CELLO is a computer program that uses BLAST to find homologous protein sequences in an in-house database annotated with the Gene Ontology database. It is used to study domains of proteins as cellular components, molecular functions, and biological processes (Yu et al., 2014). Furthermore, the presence of signal peptide was predicted by the SignalIP (5.0) server (http://www.cbs.dtu.dk/services/SignalP/) (Almagro Armenteros et al., 2019).

**Secondary and Tertiary Structure Prediction**

For secondary structure predictions of hypothetical proteins, SOPMA (https://npsa-prabi.ibcp.fr/NPSA/npsa_sopma.html), GOR1V (NPSA/npsa_gor4.html), and PSIPRED (http://www.bioinf.cs.ucl.ac.uk.) were used This tool exploits a combined approach based on the clustering process and reports biological and molecular functions based on Gene Ontology dependent similarities. Moreover, the 3D structures of selected HPs were constructed using the I-TASSER server (Zhang, 2008). The quality of protein models was evaluated via Verify3D, ERRAT, and PROCHECK programs of the SAVES server (https://saves.mbi.ucla.edu/).

**Functional characterization of HPs**

The functional characterization of hypothetical proteins was based on the conserved domain and motif analysis (Dhanyalakshmi et al., 2016). For conserved domain analysis, the NCBI CDD (www.ncbi.nlm.nih.gov/structure/cdd/docs/cdd) server and the online Pfam (https://www.pfam.xfam.org) server were used, while motif analysis, was performed using InterProScan (Ebi.ac.UK/InterPro/) Among these, the CDD compares a query sequence against position-specific score matrices using RPS-BLAST (Reverse Position Specific BLAST), resulting in the alignment of these proteins with conserved domains present in CDD (Marchler-Bauer et al., 2015). The Pfam is a protein family database with annotations and multiple sequence alignments based on Hidden Markov Models (Bateman et al., 2004). Motifs are highly conserved secondary structural elements present in protein families that define the functions of these proteins (Serçinoğlu and Ozbek, 2020). For motif analysis, InterProScan combines various methods for recognizing the protein footprints from the InterPro consortium (Paysan-Lafosse et al., 2022). After domain and motif analysis, hypothetical proteins' possible molecular and biological functions were assessed using the Argot 2.5 database (http://www.medcomp.medicina.unipd.it/Argot2-5/help.php). This tool predicts the function of the protein by its amino acid sequence in a significant manner by using a neural network approach. This combined approach is based on clustering and reports biological and molecular functions based on Gene Ontology-

dependent similarities. (Reijnders, 2022).

**Antigenicity Predictions**

VaxiJen server (http://www.jenner.ac.uk/VaxiJen9-) was used to check the antigenicity of hypothetical proteins. The potential antigenic fungal proteins were selected at the VaxiJen score greater than 0.5 (Zaharieva et al., 2019).

**Screening for Non-human Homologous HPs**

To avoid the possibility of autoimmune reactions, the drug-target proteins of *C. auris* should not be homologous to the host proteins. Therefore, only non-human homologous proteins were considered for further analysis (Khan et al., 2022). For the homology analysis, the hypothetical proteins of *C. auris*, the BLASTp analysis against humans was performed and proteins homologous to the human proteome were excluded from the analysis.

**Eessentiality Analysis of Selected Non-Human Homologous Hps**

Essential genes are those that are crucial for the survival of organisms. For essential gene analysis, the protein sequences of hypothetical proteins were subjected to the Database of Essential Genes (DEG), followed by the screening of proteins encoded by the essential genes at an e-value threshold of 0.0001 (Luo et al., 2021).

**Druggability Analysis of Essential Non-human Homologous HPs**

The essential and non-human homologous hypothetical proteins were screened against the drug bank database (https://go.drugbank.com/stats) to identify the potential drug candidates that can bind with selected drug targets. The drug bank database contains 14624 drugs, including 4243 approved drugs, 2687 biotechnology drugs, 2725 approved small molecular drugs, and 6677 experimental drugs. The matching of a hypothetical protein to a drug molecule in the drug bank database indicates its druggable properties, while the proteins not matching the existing drugs might be the novel drug targets (Murugan et al., 2020).

**Molecular Dynamics Simulations**

Molecular Dynamic (MD) simulations were performed using GROMACS software through the LINUX interface (Gomes et al., 2022). The simulations were carried out at 100ns production phase. The trajectory analysis was based on the Root Mean Square Deviations (RMSD), Root Mean Square Fluctuation (RMSF), Solubility Accessible Area, Radius

of gyration ($R_g$), and hydrogen bonds. These parameters helped in assessing the conformational stabilities of the candidate proteins.

# RESULTS

## Protein Selection and Sequence Retrieval

Initially, fifty hypothetical proteins of *Candida auris* were selected for analysis. After the sequence homology analysis by BLASTp, twelve out of fifty proteins, showing significant homology to already characterized proteins and having a percentage identity score of greater than 50 were selected for further analysis (Table 1).

## Physiochemical Characterization of HPs

The physicochemical characterization of HPs by Expasy ProtParam revealed different parameters, including molecular weight, pI, instability index, aliphatic index, and GRAVY value (Table 1). Among all twelve proteins, the KND95425.2 showed the highest molecular weight of 76374.50 Da. The isoelectric points of six HPs were above 7, while the remaining six had pI values less than 7. The instability index determines the stability of proteins, and the GRAVY value describes the hydrophobicity of proteins (Guruprasad et al., 1990; Kyte and Doolittle, 1982). The instability index value of eight HPs was greater than 40, indicating the unstable nature of these proteins, while four proteins were found to be stable with values less than 40. The GRAVY values of HPs were negative, indicating the hydrophilic nature of these proteins (Table 1).

**Table 1.** Physicochemical characterization, subcellular localization, and signal peptide prediction by using EXPASY ProtParam, CELLO2GO, and Signal IP 6.0 respectively for selected hypothetical proteins of *Candida* auris

| Sr. No | Accession Ids | Physicochemical parameters | | | | | Subcellular localization | |
|---|---|---|---|---|---|---|---|---|
| | | M.W | pI | Instability Index | Aliphatic Index | Gravy value | CELLO2GO | Signal IP 6.0 |
| 1 | KND95391.2 | 16848.78 | 9.98 | 43.18 | 81.70 | -0.387 | Mitochondrial | No |
| 2 | KND95421.2 | 9751.54 | 10.88 | 37.3 | 70.00 | -1.106 | Nuclear | No |
| 3 | KND95408.2 | 58138.2 | 6.15 | 32.11 | 107.11 | -0.365 | Plasma membrane | No |
| 4 | KND95415.2 | 75853.51 | 8.55 | 54.40 | 80.94 | -0.419 | Nuclear | No |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 5 | KND95425.2 | 76374.50 | 8.98 | 44.05 | 88.19 | -0.404 | Nuclear | No |
| 6 | KND95429.2 | 25758.28 | 6.03 | 42.18 | 89.69 | -0.539 | Nuclear | No |
| 7 | KND95462.2 | 64704.18 | 6.01 | 27.01 | 85.25 | -0.345 | Mitochondrial | No |
| 8 | KND95435.2 | 23986.09 | 5.18 | 41.05 | 74.71 | -0.764 | Nuclear | No |
| 9 | KND95434.2 | 52689.61 | 5.05 | 53.86 | 74.63 | -0.375 | Nuclear | No |
| 10 | KND95471.2 | 41927.46 | 9.50 | 46.55 | 68.27 | -0.965 | Nuclear | No |
| 11 | KND95463.2 | 15148.27 | 9.68 | 48.89 | 71.88 | -0.833 | Nuclear | No |
| 12 | KND95448.2 | 69670.96 | 6.97 | 39.76 | 89.39 | -0.196 | Mitochondrial | No |

**Subcellular localization**

Table 1 demonstrates the results of subcellular localization and signal peptide prediction of selected HPs. Three proteins, KND95462.3, KND95448.2, and KND95391.2, have scored for the mitochondrial location, while KND95408.2 was found to be associated with the plasma membrane. All remaining proteins were localized in the nuclear region (Fig. 1). SignalIP v.6.0 tool revealed that none of these proteins had signal peptides (Table 1).

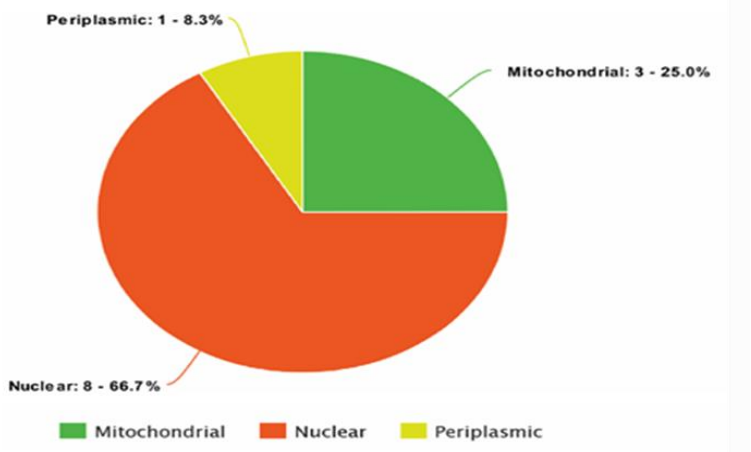


**Fig. 1.** Subcellular localization prediction of twelve HPs of *Candida auris* using CELLO2GO program

**Structure modeling and validation**

The secondary structure predictions through **SOPMA, GOR1V,** and **P SIPRED** reveal the percentage of alpha-helix, beta-sheet, and coils/loops in these proteins (Geourjon and Deleage, 1995; McGuffin et al., 2000; Sen et al., 2005). All these servers heralded that the HPs have a higher percentage of alpha helix and loops than the β-sheets (Fig. 2).

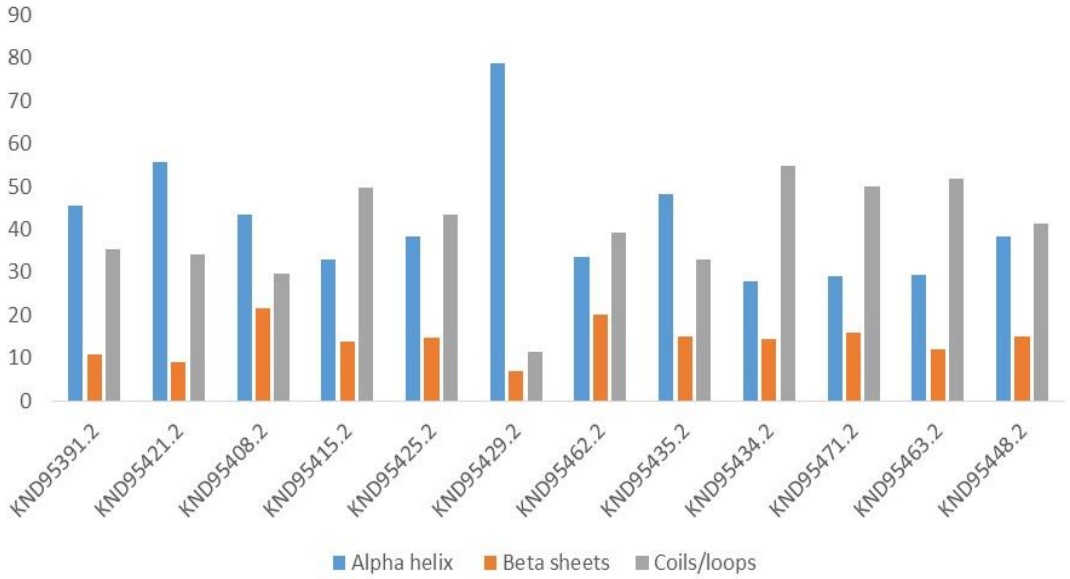


**Fig. 2.** Secondary structure prediction of HPS. The blue color indicates the alpha helix, the red color shows the percentage of beta-sheet, and the green color indicates coils/loops

**For 3D structure modeling of HPs, an I-TASSER** server was used that employs sequence-to-structure matching and then structure-to-function homology matching, thus predicting the structure and functions of the protein (Yang et al., 2015). The selection of accurate protein models was based on their C-score and TM-score. C-score is a confidence score for assessing the quality of predicted models that range from -5 to -2. Likewise, the TM score >0.5 indicates a good quality model (Steyerberg et al., 2010). Out of all protein models generated by I-TASSER, the protein model with a high C-score and TM-score was considered accurate. The 3D structures of HPs were visualized using the PyMOL server and are displayed in (Fig. 3).

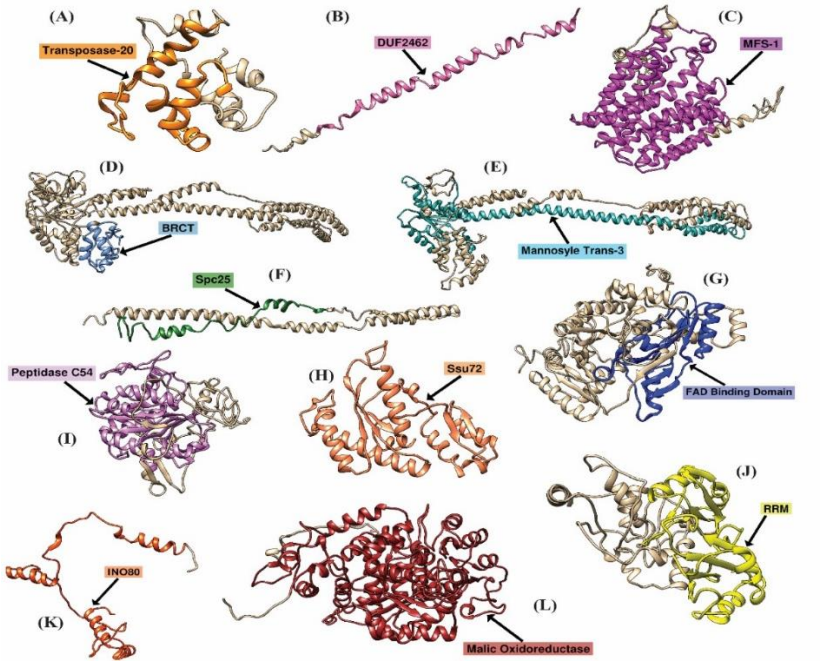


**Fig. 3.** Visualization of 3D models of HPs constructed by I-TASSER by using PYMOL and along conserved domain present in each HPs are named and highlighted in different colors in the models

After the 3D structure modeling, protein structures were validated through SAVES server version 6.0 using verify 3D, ERRAT, and PROCHECK (Dym et al., 2012; Eisenberg et al., 1997; Laskowski et al., 1993). The Verify 3D determines the compatibility of the protein model with its amino acid sequence providing a 3D-1D score. Protein models with a 3D-1D score greater than 0.2 were considered more compatible with their sequence (Eisenberg et al., 1997). ERRAT determines the overall quality factor of protein models considering all non-bonded atomic interactions. The higher ERRAT score indicates higher quality, and the generally accepted threshold for a high-quality protein model is >50 (Dym et al., 2012). Likewise, the Ramachandran plot depicts the statistical distributions of amino residues, based on their torsion angles, in the allowed and the favorable portions of the plot, giving insight into the conformation of the protein model. Based on these parameters, the protein models of all HPs were found to be accurate and of good quality (Laskowski et al., 1993) (Fig. 4).

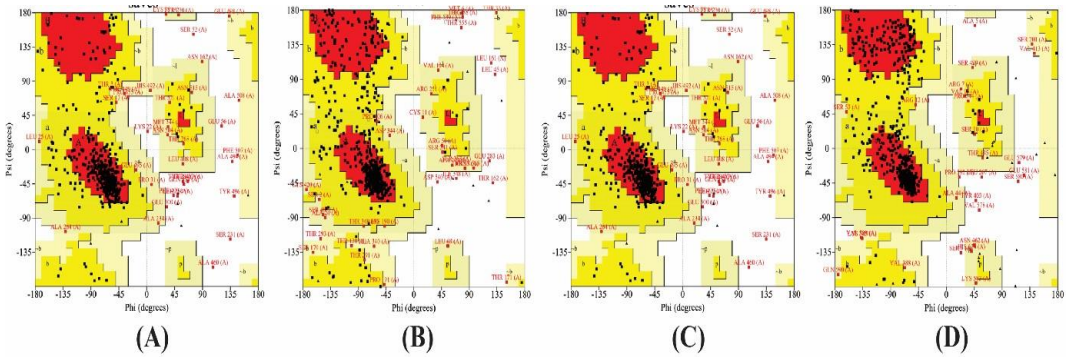


**Fig. 4.** Ramachandran Plot of 4 more compatible HPs (**A**= (KND95391.2, **B**= KND95435.2, **C**=KND95434.2, **D**=and KND95448.2)

**Conserved domain analysis and functional assessments**

NCBI conserved domain database (CDD), Pfam, and InterProScan were used to identify the conserved domains in the hypothetical proteins of *C. auris* (Bateman et al., 2004; Marchler-Bauer et al., 2015; Paysan-Lafosse et al., 2022). Several domains including, transposase-20- superfamily, DUF2462 superfamily, major facilitator superfamily (MFS), BRCT family, mannosyl transferase, chromosome segregation protein, Spc25 (Spindle –Spc25, D-lactate dehydrogenase [cytochrome], FAD linked oxidases, C-terminal domain, FAD-binding domain, RNA polymerase II subunit, Peptidase_C54, RNA recognition motif, oxaloacetate, and decarboxylating malate dehydrogenase, were identified in the HPs. The respective domains in these HPs are highlighted in their 3D structures (Fig. 3). After the conserved domain analysis, the functional annotation of hypothetical proteins was carried out using Argotv2.5, where the cut-off value of > 200 indicates the significant involvement of the protein in the predicted function (Torres et al., 2021). The functional annotation of HPs asserted their significant involvement in diverse biological pathways, including the catalysis of transposition, oxidoreductase activity, transmembrane transporter activity, DNA-directed DNA polymerase activity, nucleotide transferase activity, Flavin adenine dinucleotide activity, malate dehydrogenase, nucleic acid binding, hydrolase, and peptidase activity.

**Antigenicity prediction**

The antigenicity analysis marked six of the twelve proteins as potentially antigenic with a VaxiJen score greater than 0.5 (Fig.

5). The six proteins predicted as non-antigenic were excluded from the further analysis because these proteins could not be the drug targets.

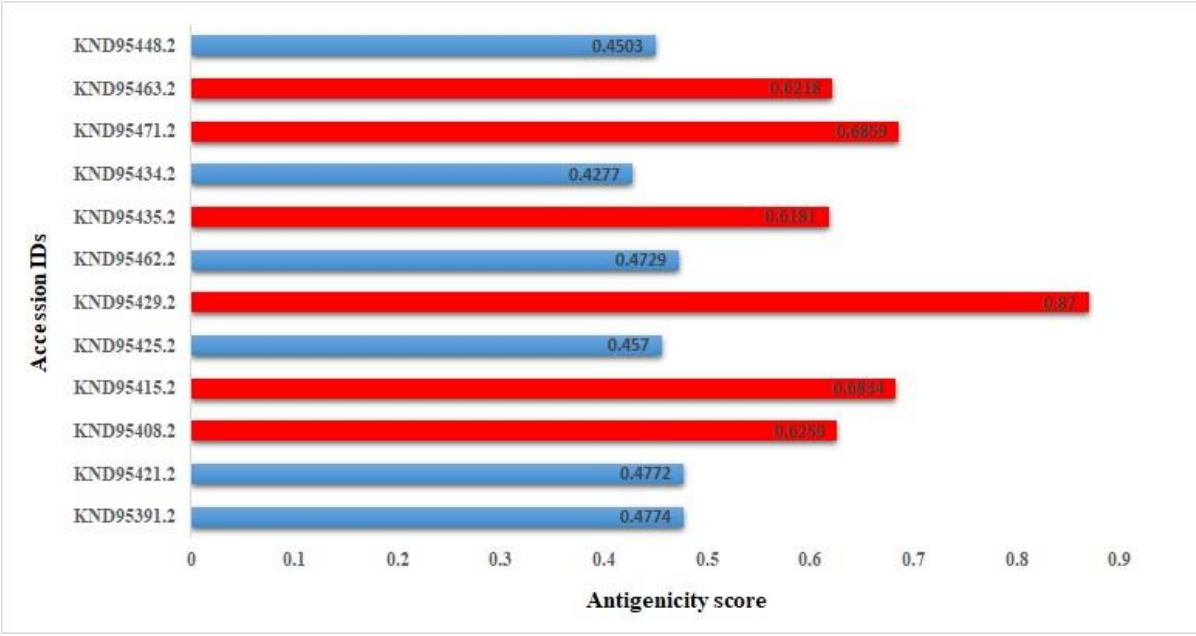


**Fig. 5.** Antigenicity prediction of HPs by using the VaxiJen server. The six proteins highlighted in red are marked as antigenic with a VaxiJen score greater than 0.5, while the blue-labeled proteins are non-antigenic

**Screening of non-human homologous proteins**

A notable feature of the drug candidate is that it should not be homologous to the host proteome to avoid cross-reactivity in hosts (Pourhajibagher and Bahador, 2016). Therefore, the homology analysis of selected immunogenic proteins against the human proteome was performed using BLASTp. The results revealed that three proteins have significant homology with human proteome (Fig. 6). The three human homologous proteins were excluded from analysis, and three non-human homologous HPs were further analyzed to probe the potential drug targets.

**Essentiality analysis of selected non-human homologous HPs**

After the homology analysis with the human proteome, the essentiality analysis reveals that three non-human homologous proteins are also crucial for the survival of *Candida auris* thus, these three HPs were selected as the drug targets for *C. auris* and were subjected to druggability analysis (Fig. 6).

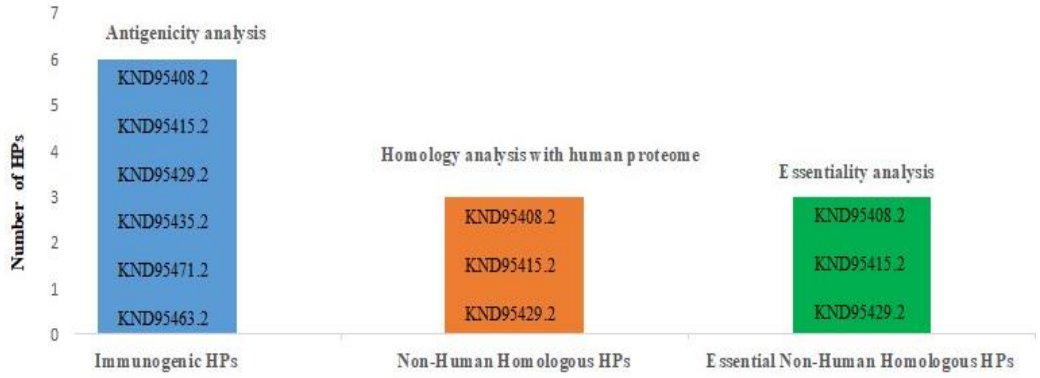


**Fig. 6:** The immunogenic, non-human homologous and essential hypothetical proteins of *Candida auris*

**Druggability Analysis of Essential Non-Homologous Hypothetical Proteins**

The druggability analysis of three selected non-human homologous and essential proteins marked them as novel drug targets having no affinity with the already available drug molecules (Table 2).

**Table 2:** Druggability analysis of essential non-homologous hypothetical proteins by Drug Bank.

| Sr.No | Protein Accession ID | Druggability analysis |
|---|---|---|
| 1 | KND95408.2 | Novel |
| 2 | KND95415.2 | Novel |
| 3 | KND95429.2 | Novel |

**Molecular dynamics simulations**

MD simulations of the three selected drug target proteins showed in (Fig. 7). The combined analysis of Root Mean Square Deviations (RMSD), Root Mean Square Fluctuation (RMSF), Surface Accessible Area, Radius of gyration (Rg), and the hydrogen bonds shows the stable conformations of three proteins with the different fluctuations over time. The protein KND95408.2 remains compact with the least fluctuations, while the proteins KND95415.2 and KND95429.2 show marginal deviations without undergoing any drastic changes in confirmations.

MD simulations depict that the selected candidates are the best potent drug targets depending on

the nature of the proteins determined by multiple sequence alignment after detecting their homologs. As KND95408.2 is a membrane transporter having a compact nature it showed less variation and fluctuations in its structure and remained stable over time. While the protein, KND95415.2 has a significant role in DNA damage repair, cell cycle regulation, and checkpoint control so, the fluctuation and conformational changes in its structure depict its nature and give confidence about the promising results. Respectively, KND95429.2 proteins despite their small size have high fluctuation and conformational changes that synchronize with their functional characterization as it has significant importance in cell division, maintenance of genome stability, and repair of double-stranded DNA breaks. The nature of proteins along with their function co-relate with the conformational changes detected by MD simulations.

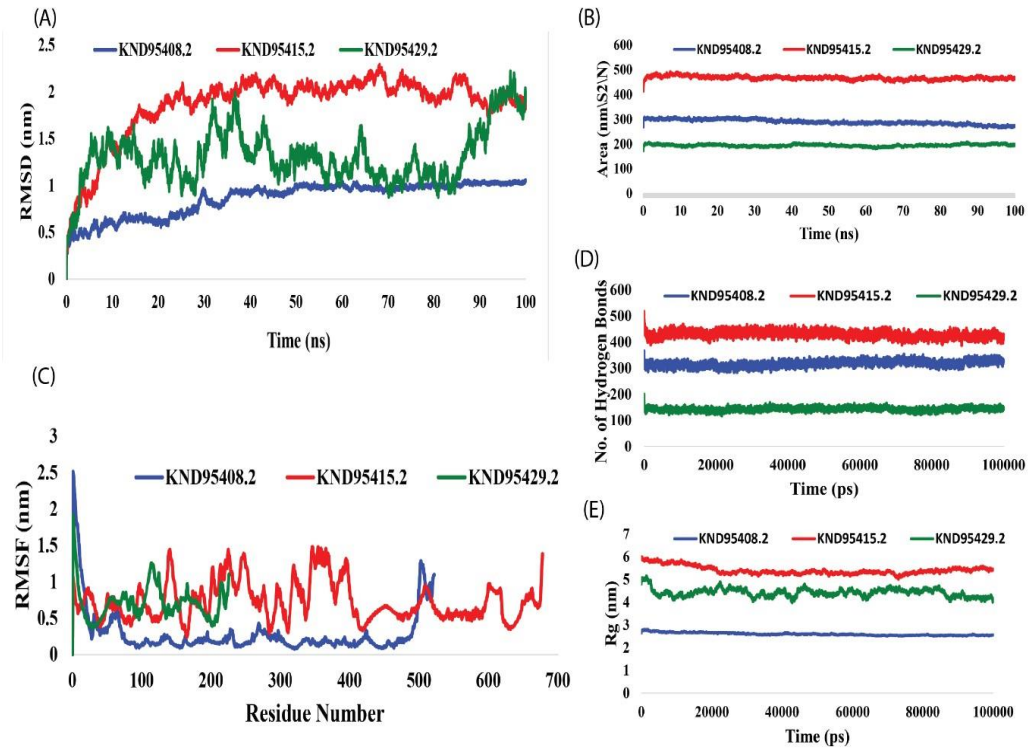


Fig. 7: Simulation analysis of hypothetical proteins of *Candida auris* using GROMACS. (A) RMSD at 100ns shows significant conformational changes in Red and Green proteins, while Blue remains more stable (B) Solubility Accessible Area reflects protein stability with hydrophobic core exposure (C) RMSF at 100ns indicates higher flexibility in Red and Green proteins, with Blue being less flexible (D) Number of Hydrogen Bonds indicates greater compactness in the larger Red protein (E)

Radius of Gyration shows Green protein is short and flexible with a large radius, while Blue is compact with a small radius. Red protein is the largest and most flexible

## DISCUSSION

*Candida auris* cause infections, collectively known as candidiasis (Du et al., 2020), and due to its resistance to antifungal drugs and high mortality rate, the Center for Disease Control and Prevention (CDC) has declared this pathogen a superbug (D'Ambra, 2019). The emergence of drug resistance in *C. auris* demands the exploration of new treatment options to combat this pathogen (Jain et al., 2022). The structural and functional characterization of hypothetical proteins can pave the way, to identify new potential drug targets facilitating drug development (Prabhu et al., 2020). Many novel drug targets have been identified for various pathogens from their uncharacterized proteins (Ahamed et al., 2021; Araújo et al., 2020; Hafsa et al., 2022; Pranavathiyani et al., 2020). Since a large portion of the proteome of *C. auris* remains structurally and functionally uncharacterized, the structural and functional annotations of these proteins would help to determine the novel therapeutic targets of this pathogen (Rossato and Colombo, 2018). The present study employs various computational algorithms to characterize hypothetical proteins from *C. auris* to discover novel therapeutic targets from these uncharacterized proteins. Initially, fifty hypothetical proteins of *C. auris* were retrieved from the NCBI database. The hypothetical proteins showing significant sequence similarity to the already characterized proteins were screened through BLASTp analysis because the structurally and functionally characterized proteins having homology with the uncharacterized proteins might help to infer the structure and function of hypothetical proteins (Ahmed et al., 2018). Out of fifty hypothetical proteins, twelve homolog proteins were considered for further analysis, whose physicochemical characterization provides insights into the chemical nature, stability, and molecular attributes (Panda and Chandra, 2012). The Expasy ProtParam tool revealed that one hypothetical protein (KND95425.2) has the highest molecular weight of 76 kDa, and the GRAVY values of all twelve hypothetical proteins were negative, indicating the non-polar nature of these proteins (Table 1). As the function of a protein is directly or indirectly related to its subcellular localization, which indicates that two proteins are

mitochondrial, one protein is the plasma membrane protein, and all remaining proteins are nuclear. None of these hypothetical proteins shows the presence of signal peptide as predicted by the SignalIP (6.0) server (Table 1 and Fig. 1). Structure modeling of the proteins is crucial to evaluate their molecular and biological functions in cells, thus assisting in drug target identification (Zhang, 2009). The secondary structural modeling of HPs reveals the significant presence of alpha-helix and coils/loops, then beta-sheets (Fig. 2). After the 2D structure predictions, the tertiary structures of proteins were modeled using I-TASSER, followed by the quality assessment of protein structures. The protein structure validation through these programs revealed that the selected models have minimum steric hindrance with most of their amino acid residues in favorable and allowed regions of the Ramachandran plot (Fig. 4). The functional annotation of the hypothetical protein is a prerequisite for a better understanding of biological processes and finding potential drug targets from them (Loewenstein et al., 2009). NCBI conserved domain database, Pfam, and InterProScan revealed the presence of evolutionarily conserved domains in these proteins; moreover, the Argot 2.5 server revealed the involvement of all hypothetical proteins in the crucial biological pathways (Fig. 3).

The pathogenic proteins of microbes play a pivotal role in the initial interaction between the pathogen and the host. Consequently, these proteins are frequently prioritized as suitable drug targets. In addition to their immunogenic nature, a drug target protein must also be crucial for the survival of the pathogen. Furthermore, these proteins should not have similarities with the host proteins, which otherwise may cause the cross reactivity. Considering these points the selected hypothetical proteins were analyzed for their immunogenic nature, homology with the human proteome, and their essentiality in the survival of pathogen (Qureshi et al., 2021; Sauna et al., 2020). The immunogenic predictions through the VaxiJen server revealed six proteins as antigenic (Fig. 5). The homology analysis marked three among the six antigenic proteins as non-human homologous, and essentiality analysis indicated that three non-human homologous proteins are also essential for the survival

of *C. auris* (Fig. 6). Therefore, the three immunogenic, non-human homologous, and essential proteins were declared as the potential drug target proteins for *C. auris* (Table 2). Finally, the similarity search against the Drug Bank database showed that all three immunogenic, essential, and non-human homologous hypothetical proteins have no significant similarity with compounds in the Drug Bank database; therefore, these proteins are novel drug targets (Table 2). The Molecular dynamic simulations show the thermodynamically stable conformation of the selected hypothetical proteins, thereby validating their suitability as promising drug targets (Fig. 7). Based on this *in silico* screening, characterization, and simulation analysis KND95408.2, KND95415.2, and KND95429.2 are potential drug targets, but they need further wet-lab characterization to confirm their suitability as drug targets.

## CONCLUSION

The structural and functional characterization of hypothetical proteins is crucial for understanding their role in biochemical/physiological pathways and screening novel therapeutic targets. The present study employs an *in-silico* approach to identify novel drug targets from hypothetical proteins of *Candida auris* that are found to be essential for this pathogen and are non-human homologous. However, these proteins need further wet lab characterization before their use in drug designing.

### ACKNOWLEDGMENT

Authors are thankful for the online available open access databases which were used in this research. The research was self-supported by the authors.

### ETHICAL STATEMENT

This study is a computational work. No humans or animals were used during the study. Therefore, ethical approval was not necessary.

### CONFLICT OF INTEREST

Authors declare no conflict of interests.

### DATA AVAILABILITY

The data that supports the findings of this study are available on request by the corresponding author.

## REFERENCES

1. Abbasi, B. A., Dharan, A., Mishra, A., Saraf, D., Ahamad, I., Suravajhala, P., and Valadi, J. (2022). In

silico characterization of uncharacterized proteins from multiple strains of *Clostridium difficile*. *Frontiers in Genetics*, 13, 878012.

2. Ahamed, N. A., Panneerselvam, A., Arif, I. A., Abuthakir, M. H. S., Jeyam, M., Ambikapathy, V., Health, P. (2021). Identification of potential drug targets in human pathogen *Bacillus cereus* and insight for finding inhibitor through subtractive proteome and molecular docking studies. *Journal of Infection and Public Health*, 14(1), 160-168.
3. Ahmed, M. S., Shahjaman, M., Kabir, E., and Kamruzzaman, M. J. B. (2018). Structure modeling to function prediction of uncharacterized human protein C15orf41. *BMC Bioinformatics*, 14(5), 206.
4. Almagro Armenteros, J. J., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., ... Nielsen, H. J. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnology*, 37(4), 420-423.
5. Almeida, F., Rodrigues, M. L., and Coelho, C. (2019). The still underestimated problem of fungal diseases worldwide. *Frontiers in Microbiology*, 10, 214.
6. Araújo, C. L., Blanco, I., Souza, L., Tiwari, S., Pereira, L. C., Ghosh, P., Folador, A. (2020). In silico functional prediction of hypothetical proteins from the core genome of *Corynebacterium pseudotuberculosis* biovar ovis. *PeerJ*, 8, e9643.
7. Bandara, H., and Samaranayake, L. J. (2022). Emerging strategies for

environmental decontamination of the nosocomial fungal pathogen *Candida auris*. *Journal of Medical Microbiology*, 71(6), 001548.

8. Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., Sonnhammer, E. L. (2004). The Pfam protein families' database. *Nucleic Acids Research*, 32(suppl_1), D138-D141.
9. Büyüktuna, S. A., Hasbek, M., Elaldı, N., Gözel, M. G., Çelik, C., Engin, A., Bakır, M. (2019). Epidemiological analysis of nosocomial *Candida* infections: Experience of a university hospital. *Clinical Microbiology and Infection*, 41(2), 318-327.
10. Cândido, E. d. S., Affonseca, F., Cardoso, M. H., and Franco, O. L. (2020). Echinocandins as biotechnological tools for treating *Candida auris* infections. *Journal of Fungi*, 6(3), 185.
11. Chirgadze, Y., Boshkova, E., Kargatov, A., Chirgadze, N., and Dynamics. (2022). Functional identification of 'hypothetical protein' structures with unknown function. *Journal of Biomolecular Structure and Dynamics*, 1-5.
12. D'Ambra, M. (2019). CDC alert for hospital infections due to *Candida auris* multiresistant. *Clinical Translational Gastroenterology*, 2, 150-152.
13. Dahiya, S., Chhillar, A. K., Sharma, N., Choudhary, P., Punia, A., Balhara, M., Parmar, V. S. (2020). *Candida auris* and nosocomial infection. *Current Drug Targets*, 21(4), 365-373.

14. Dhanyalakshmi, K., Naika, M. B., Sajeevan, R., Mathew, O. K., Shafi, K. M., Sowdhamini, R., and N. Nataraja, K. (2016). An approach to function annotation for Proteins of Unknown Function (PUFs) in the transcriptome of Indian mulberry. *PLoS ONE*, 11(3), e0151323.
15. Du, H., Bing, J., Hu, T., Ennis, C. L., Nobile, C. J., and Huang, G. (2020). *Candida auris*: Epidemiology, biology, antifungal resistance, and virulence. *PLoS Pathogens*, 16(10), e1008921.
16. Duvaud, S., Gabella, C., Lisacek, F., Stockinger, H., Ioannidis, V., and Durinx, C. (2021). Expasy, the Swiss Bioinformatics Resource Portal, as designed by its users. *Nucleic Acids Research*, 49(W1), W216-W227.
17. Dym, O., Eisenberg, D., and Yeates, T. (2012). ERRAT.
18. Eisenberg, D., Lüthy, R., and Bowie, J. U. (1997). [20] VERIFY3D: assessment of protein models with three-dimensional profiles. In Methods in enzymology (Vol. 277, pp. 396-404). Elsevier.
19. Fang, J., Huang, B., and Ding, Z. (2021). Efficacy of antifungal drugs in the treatment of oral candidiasis: A Bayesian network meta-analysis. *Journal of Pharmaceutical Design*, 125(2), 257-265.
20. Geourjon, C., and Deleage, G. (1995). SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Bioinformatics*, 11(6), 681-684.

21. Gomes, D. E. B., Gomes, P. S., and Bernardi, R. C. (2022). QwikMD 2.0: bridging the gap between sequence, structure, and protein function. *Biophysical Journal*, 121(3), 132a.
22. Guruprasad, K., Reddy, B. B., and Pandit, M. W. (1990). 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 Engineering*, 4(2), 155-161.
23. Hafsa, U., Chuwdhury, G., Hasan, M. K., Ahsan, T., and Moni, M. A. (2022). An in silico approach towards identification of novel drug targets in *Klebsiella oxytoca*. *Infection, Genetics and Evolution*, 31, 100998.
24. Horton, M. V., and Nett, J. E. (2020). *Candida auris* infection and biofilm formation: going beyond the surface. *Clinical Candida Mycology Research*, 7(3), 51-56.
25. Jain, M., Jain, A., Khare, B., Jain, D. K., Khan, R., and Jain, D. (2022). An update on the recent emergence of *Candida auris*. *American Journal of Dermatological and Health Sciences*, 2(1), 14-19.
26. Khan, K., Jalal, K., Khan, A., Al-Harrasi, A., and Uddin, R. (2022). Comparative metabolic pathways analysis and subtractive genomics profiling to prioritize potential drug targets against *Streptococcus pneumoniae*. *Frontiers in Microbiology*, *12*, 4384.
27. Kyte, J., and Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *Journal of*

*Molecular Biology*, *157*(1), 105-132.

28. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993). PROCHECK: A program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, *26*(2), 283-291.
29. Loewenstein, Y., Raimondo, D., Redfern, O. C., Watson, J., Frishman, D., Linial, M., ... Tramontano, A. (2009). Protein function annotation by homology-based inference. *Genome Biology*, *10*(2), 1-8.
30. Luo, H., Lin, Y., Liu, T., Lai, F.-L., Zhang, C.-T., Gao, F., and Zhang, R. (2021). DEG 15, an update of the Database of Essential Genes that includes built-in analysis tools. *Nucleic Acids Research*, *49*(D1), D677-D686.
31. Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Geer, L. Y., Hurwitz, D. I. (2015). CDD: NCBI's conserved domain database. *Nucleic Acids Research*, *43*(D1), D222-D226.
32. McGuffin, L. J., Bryson, K., and Jones, D. T. (2000). The PSIPRED protein structure prediction server. *Bioinformatics*, *16*(4), 404-405.
33. Murugan, N. A., Kumar, S., Jeyakanthan, J., and Srivastava, V. (2020). Searching for target-specific and multi-targeting organics for COVID-19 in the DrugBank database with a double scoring approach. *Scientific Reports*, *10*(1), 1-16.
34. Omeershffudin, U. N. M., and Kumar, S. (2019). In silico approach for mining of potential drug targets

from hypothetical proteins of bacterial proteome. *International Journal of Microbiology Open Access*, *4*(4), 145-152.

35. Panda, S., and Chandra, G. (2012). Physicochemical characterization and functional analysis of some snake venom toxin proteins and related non-toxin proteins of other chordates. *Bioinformatics*, *8*(18), 891.
36. Paysan-Lafosse, T., Blum, M., Chuguransky, S., Grego, T., Pinto, B. L., Salazar, G. A., Colwell, L. J. (2022). InterPro in 2022. *Nucleic Acids Research*.
37. Pourhajibagher, M., and Bahador, A. (2016). Designing and in silico analysis of PorB protein from *Chlamydia trachomatis* for developing a vaccine candidate. *Drug Research*, *66*(09), 479-483.
38. Prabhu, D., Rajamanikandan, S., Anusha, S., Chowdary, M. S., Veerapandiyan, M., and Jeyakanthan, J. (2020). In silico functional annotation and characterization of hypothetical proteins from *Serratia marcescens* FGI94. *BioBiotechnology*, *47*(4), 319-331.
39. Pranavathiyani, G., Prava, J., Rajeev, A. C., and Pan, A. (2020). Novel target exploration from hypothetical proteins of *Klebsiella pneumoniae* MGH 78578 reveals a protein involved in host-pathogen interaction. *Frontiers in Cell and Infection Microbiology*, *109*.
40. Qureshi, N. A., Bakhtiar, S. M., Faheem, M., Shah, M., Bari, A., Mahmood, H. M., ... Jamal, S. B. (2021). Genome-based drug target identification in human pathogen *Streptococcus*

*gallolyticus*. *Frontiers in Genetics*, *12*, 564056.

41. Ré, A. C. S., Martins, J. F., Cunha-Filho, M., Gelfuso, G. M., Aires, C. P., and Gratieri, T. (2021). New perspectives on the topical management of recurrent candidiasis. *Drug Development and Research*, *11*(4), 1568-1585.
42. Reijnders, M. J. (2022). Wei2GO: Weighted sequence similarity-based protein function prediction. *Protein Science*, *10*, e12931.
43. Rossato, L., and Colombo, A. L. (2018). *Candida auris*: What have we learned about its mechanisms of pathogenicity? *Frontiers in Microbiology*, *9*, 3081.
44. Sauna, Z. E., Richards, S. M., Maillere, B., Jury, E. C., and Rosenberg, A. S. (2020). Immunogenicity of proteins used as therapeutics. *Frontiers in Immunology*, *11*, 614856.
45. Schoch, C. L., Ciufo, S., Domrachev, M., Hotton, C. L., Kannan, S., Khovanskaya, R., Robbertse, B. (2020). NCBI Taxonomy: A comprehensive update on curation, resources, and tools. *Database*.
46. Segrelles-Calvo, G., de S Araújo, G. R., Llopis-Pastor, E., Carrillo, J., Hernández-Hernández, M., Rey, L., Zamarro, C. (2021). *Candida* spp. co-infection in COVID-19 patients with severe pneumonia: Prevalence study and associated risk factors. *Respiratory Medicine*, *188*, 106619.
47. Sen, T. Z., Jernigan, R. L., Garnier, J., and Kloczkowski, A. (2005). GOR V server for protein secondary structure

prediction. *Bioinformatics*, *21*(11), 2787-2788.

48. Serçinoğlu, O., and Ozbek, P. (2020). Sequence-structure-function relationships in class I MHC: A local frustration perspective. *PLoS One*, *15*(5), e0232849.
49. Shamsinejad, F. S., Zafari, Z., and Zafari, Z. (2022). Prediction of potential drug targets and vaccine candidates against antibiotic-resistant *Pseudomonas aeruginosa*. *International Journal of Pharmaceutical Research and Therapeutics*, *28*(6), 1-11.
50. Steyerberg, E. W., Vickers, A. J., Cook, N. R., Gerds, T., Gonen, M., Obuchowski, N., Kattan, M. W. (2010). Assessing the performance of prediction models: A framework for some traditional and novel measures. *Epidemiology*, *21*(1), 128.
51. Thomas-Rüddel, D. O., Schlattmann, P., Pletz, M., Kurzai, O., and Bloos, F. (2022). Risk factors for invasive *Candida* infection in critically ill patients: A systematic review and meta-analysis. *Clinical Microbiology*, *161*(2), 345-355.
52. Torres, M., Yang, H., Romero, A. E., and Paccanaro, A. (2021). Protein function prediction for newly sequenced organisms. *Nature Microbiology Insights*, *3*(12), 1050-1060.
53. Vila, T., Sultan, A. S., Montelongo-Jauregui, D., and Jabra-Rizk, M. A. (2020). *Candida auris*: A fungus with an identity crisis. *Pathogens and Disease*, *78*(4), ftaa034.
54. Wang, R., and Chen, L. (2022). Identification of

human protein subcellular location with multiple networks. *Cellular Physiology*, *19*(4), 344-356.

55. Wu, B., Zhang, H., Lin, L., Wang, H., Gao, Y., Zhao, L., Gu, L. (2019). A similarity searching system for biological phenotype images using deep convolutional encoder-decoder architecture. *Computational Biology*, *14*(7), 628-639.

56. Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., and Zhang, Y. (2015). The I-TASSER Suite: Protein structure and function prediction. *Nature Methods*, *12*(1), 7-8.

57. Yu, C.-S., Cheng, C.-W., Su, W.-C., Chang, K.-C., Huang, S.-W., Hwang, J.-K., and Lu, C.-H. (2014). CELLO2GO: A web server for protein subCELlular LOcalization prediction with functional gene ontology annotation. *PLOS One*, *9*(6), e99368.

58. Zaharieva, N., Dimitrov, I., Flower, D. R., and Doytchinova, I. (2019). VaxiJen dataset of bacterial immunogens: An update. *Current Computer-Aided Drug Design*, *15*(5), 398-400. Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*, *9*(1), 1-8.

59. Zhang, Y. (2009). Protein structure prediction: When is it useful? *Current Opinion in Structural Biology*, *19*(2), 145-155.