



Original Research

Structural and Functional Analyses of Hypothetical Protein of *Saccharomyces cerevisiae* Reveals Cytochrome C Oxidase Subunit-I Protein: In Silico Approaches

Md. Nafij Mashrur^{1,2}, Md. Rakibul Hasan¹, Sadia Sultana^{2,5}, Mitnur Bobby^{3,5}, Mannalal das^{4,5}, Sahil Abrar^{2,5}, Mst Rubaiat Nazneen Akhand^{2,5}

¹Department of Biochemistry and Molecular Biology, Gono University, Dhaka 1344, Bangladesh.

²Department of Biochemistry and Chemistry, Sylhet Agricultural University, Sylhet 3100, Bangladesh.

³Department of Pharmaceuticals and Industrial Biotechnology, Sylhet Agricultural University, Sylhet 3100, Bangladesh.

⁴Department of Animal and Fish Biotechnology, Sylhet Agricultural University, Sylhet 3100, Bangladesh.

⁵Faculty of Biotechnology and Genetic Engineering, Sylhet Agricultural University, Sylhet 3100, Bangladesh.

Article info

Received: October 21, 2025

Revised: November 07, 2025

Accepted: November 12, 2025

Published: November 21, 2025

*Corresponding author

Department of Biochemistry and Molecular Biology, Gono Bishwabidyalay (University), Nolam, Savar, Dhaka 1344, Bangladesh
E-mail: mn.mashrur@gmail.com

ORCID ID: <https://orcid.org/0009-0007-3167-0903>

Keywords: *Saccharomyces cerevisiae*, mitochondrial DNA (mtDNA), hypothetical protein, cytochrome c oxidase-like subunit I, molecular docking.

Cite this article: Mashrur MN, Hasan MR, Sultana S, Bobby M, Das M, Abrar S, Akhand MRN. Structural and functional analyses of hypothetical protein of *Saccharomyces cerevisiae* reveals cytochrome c oxidase subunit I protein: In silico approaches. J Biosci Public Health. 2025;1(3):79-97. doi:10.5455/JBPH.2025.16

ABSTRACT

The structural and functional investigation of a hypothetical protein HP (KZV07215.1) from *Saccharomyces cerevisiae* (SC) has shed light on how fungal adaptation to unusual conditions and metabolic modifications occur. In this study, we performed a comprehensive in silico characterization of a mitochondrial hypothetical protein (KZV07215.1), encoded by the WN66_06778 gene, to elucidate its putative structure and function. Physicochemical profiling revealed a moderately acidic, hydrophobic, and highly stable 58.7-kDa protein with an isoelectric point of 6.55. According to the results of functional annotating programs like InterProScan, NCBI-CDD Search, and Pfam, the desired product is a cytochrome c oxidase (COX) protein of the respiratory system. Its 3D configuration was determined using the homology modeling technique by I-TASSER (TM-score estimation of 0.98 ± 0.05). Following the GalaxyRefine, a stable 3D structure was confirmed by PROCHECK, QMEAN (value of -3.36), QMEANDisCo Global (score of 0.77 ± 0.05), and ERRAT (score of 92.381). FTSite confirms the site of activity of the hypothesized structure. Molecular docking revealed substantial binding affinities with HEA compounds, yielding docking scores of -159.90 kcal/mol (HP-HEA III), ensuring robust and stable interactions. These findings strongly support the annotation of KZV07215.1 as a COX-I-like protein involved in electron transport chain complex IV activity. This work highlights the value of computational pipelines for functional annotation of uncharacterized mitochondrial proteins and suggests potential implications for understanding respiratory regulation and bioenergetic disorders.

1. INTRODUCTION

Saccharomyces cerevisiae (SC) is a unicellular, yellow-green, spherical yeast belonging to the fungi kingdom, having 16 chromosomes and a total genomic DNA of 12068 kilobases (kb). It is a tiny, single-cell fungus that can be readily cultured. It takes quick production, around 1.25-2 hours, to double at a temperature of 30°C, and allows the affordable maintenance of different variants [1]. As a facultative anaerobe, SC survives in both oxygen-rich and oxygen-limited environments, deriving energy through mitochondrial oxidative phosphorylation and the electron transport chain, as well as through fermentative pathways. Because of its metabolic versatility and genetic tractability, SC has long served as a cornerstone in mitochondrial genetics. Pioneering studies on respiratory-deficient “petite” mutants established cytoplasmic inheritance [2], and SC became the organism in which mitochondrial genes [3] and their mosaic intronic architecture [4,5] were first described. It was also the source of the earliest sequenced mitochondrial gene [6, 7]. Publication of the complete yeast genome in 1996 marked a turning point in functional genomics, revealing that nearly half (56%) of annotated genes lacked experimentally defined roles despite decades of biochemical and genetic research [8-11]. This unexpected knowledge gap stimulated large-scale functional genomics initiatives aimed at characterizing previously unknown proteins [12]. Even today, more than half of the predicted proteins in *S. cerevisiae* remain functionally uncharacterized, reflecting substantial gaps in our understanding of mitochondrial biology and proteome complexity. Experimental elucidation of such proteins is challenging due to technical, temporal, and financial constraints. Consequently, integrative computational approaches provide a powerful and efficient alternative for predicting the structural and functional properties of hypothetical mitochondrial proteins [13].

The respiratory chain's terminal enzyme, cytochrome C oxidase (COX), catalyzes the coordinated movement of ferrocycytochrome C's multiple (four) electrons towards oxygen molecules. Again, pumping protons also moves from the mitochondrial matrix to the mitochondrial inner cytosolic portion [14]. The catalytically active holoenzyme is made up of 4 reactive metallic cores (Heme A, Heme A₃, Cu_A, and Cu_B) comprised within a multi-subunit complex. The core of the binuclear reaction is made up of a bridge in-between the completely oxidative relaxation stage of Cu_B and Heme A₃ [15]. Many polypeptide subunits make up the protein matrix that envelops the metal centers. The initial subunit binds the reductive cores of the proteins Heme A, Heme A₃, and Cu_B in COXs originate from prokaryotic systems along with eukaryotic cell types. In addition to binding Cu_A, subunit II also engages in the binding of cytochrome c, most likely at a location near Cu_A [16]. The functioning of components I and II's proton transport may be modulated by subunit III [17]. The subunits that are responsible for the making up an active COX molecule are substantially more than three in eukaryotes. The mitochondrial genome codes for the three main subunits (I-III) of the eukaryotic COX complex. They possess basic sequence homology to those of three subunits of bacterial cytochrome c oxidases [18]. The remaining subunits are all products of the nuclear genome. SC serves as a versatile eukaryotic model for researchers to investigate mitochondrial biogenesis and the assembly of respiratory complexes. It can endure the absence of oxidative phosphorylation by using fermentation, which distinguishes it from most organisms. This enables researchers to examine alterations that would be lethal in other creatures.

Furthermore, both its nuclear and mitochondrial genomes are amenable to genetic modification, making it an excellent model for investigating mitochondrial gene function [19, 20].

This work aims to define a potential mitochondrial protein of *S. cerevisiae* (accession no. KZV07215.1) using a comprehensive in silico workflow integrating physicochemical profiling, domain annotation, evolutionary analysis, secondary and tertiary structure prediction, and molecular docking. By elucidating its putative relationship to COX-I, this work provides insights into its potential functional role in the yeast electron transport chain and lays the foundation for future experimental validation.

2. MATERIALS AND METHODS

2.1. Sequence Retrieval

The NCBI (accessed February 2025) (National Center for Biotechnology Information) Protein database was searched for the keyword "Hypothetical proteins AND mitochondrion AND *Saccharomyces cerevisiae*," and the 766 residues of HP of SC were retrieved. From the list of hits, we randomly selected a hypothetical protein with accession number KZV07215.1 (GI|1023939894|) and retrieved its FASTA sequence for further investigation. A peptide search based on sequence was also carried out to ascertain whether the protein is unnecessary or not, using UniProt peptide search [21].

2.2. Physicochemical Properties Analysis

The ProtParam software in the online ExPASy server (accessed February 2025) was employed in order to evaluate preferred HPs chemically mediated as well as physical features. The analyzer yields theoretical measurements including weight of molecule, amino acid composition, theoretical pI value, count of both positive and negative residues, extinction coefficient, GRAVY score (grand average of hydropathicity), etc [22].

2.3. Functional Domain Annotation

The HP was functionally annotated to make its functions clear. Several freely accessible databases and software, including the NCBI conserved domains database (CDD v3.21, updated 2024) [23], InterProScan v5.57-90.0 (2024 release) [24], and SUPERFAMILY v1.75 [25], were implemented for proper screening of HP's functional domain and conserver. In every situation, the default parameters were used. The proteins are categorized using the screened-out domains found in the HP. The domains are found using these databases and additional bioinformatics techniques [26].

2.4. Phylogenetic Analysis and Visualization

The UniProtKB (accessed February 2025) [27] was implemented, looking for similar sequences with the examined HP. Using UniProt's BLAST technique [28] We looked for similarities in a database of unique proteins. Initially, a total of 42 reviewed homologous protein sequences were identified from the UniProt database. Among them, the first 24 reviewed sequences were downloaded. They were all thought to serve the same function.

The multiple sequence alignment (MSA) was then done by utilizing the MAFFT v7.526 (April 2024) tool [29]. Using homologous sequence alignment, a phylogenetic tree was generated with PhyML v3.0 (release 2024) to highlight the evolutionary segregation of the related proteins [30]. The tree was shown by using the iTOL v6 (accessed February 2025) server [31].

2.5. Secondary Structure Prediction

The prediction of the two-dimensional (2D) structure of HP was shown by the SOPMA servers [32, 33] and PSIPRED v4.0 (2024 release) [34, 35]. With the use of "DATABASE.DSSP," SOPMA makes secondary structure predictions for proteins, whereas PSIPRED utilizes the PSI-BLAST algorithm with feed-forward neural networks [34, 35]. In both cases, the prediction of the 2D structure was performed using the FASTA sequence of HP.

2.6. Tertiary Structure Prediction

The prediction of three-dimensional (3D) structure was carried out by employing the I-TASSER (version 5.1, 2023) servers [36–38]. For homology modeling, we used the default values for all the variables of I-TASSER [39]. The anticipated three-dimensional structures of the HP were improved, and then the YASARA v25.9.17 energy minimization server was implemented to reduce their energy [40]. After that, the enhanced 3D structures were further improved using GalaxyRefine [41]. GalaxyRefine produces several potential structures; the highest-performing and highest-quality ones are hand-selected. The structures of the HP were then converted into three-dimensional forms using PyMOL v2.5.2 [42, 43].

2.7. Model Quality Assessment

SAVES v6.0 (2024 release; accessed March 2025) server's ERRAT [44] modules along with PROCHECK [45] were accessible for the assessment of HP's energy-minimization and ideal 3D structure. The 3D model of HP was validated further with the SWISS-MODEL (release 2024) Structural Assessment Tool, which was created by the Swiss Institute of Bioinformatics (SIB) [46, 47]. Finally, the highest standard model was chosen to support further investigation.

2.8. Subcellular Localization Prediction

The optimal functioning of peptides is influenced by their spatial surroundings, which control their biological networks and patterns of interaction [48]. In this particular context, several servers anticipated the HP's subcellular location, including Euk-mPLoc v2.0 [49], TargetP-v2.0 [50], and MitPred v2.0 (accessed March 2025) servers [51].

2.9. Binding Site Prediction and Molecular Docking

The FTSite (2012; accessed March 2025) server was chosen for the determination of the active site and residues of the HP [52]. Quick residue-level annotations are achievable only when the structure and sequence of a protein have a relationship [53]. To ascertain the binding potency of the HP with Heme-A (HEA), molecular docking was conducted utilizing the HDOCK (2024; accessed April 2025) online service [54, 55]. The ligand HEA's PDB files were acquired from the RCSB PDB [56], and the I-TASSER-generated HP was used as the receptor. The sets of complexes with high negative docking values were selected and retrieved, indicating excellent docking. Version v2.5.2 of PyMOL was used to visualize the obtained result from the HDOCK server [42, 43].

3. RESULTS

3.1. Retrieval of HP Sequence

The protein database at NCBI was randomly queried, resulting in the HP WN66 06778, which is SC’s HP. Once this sequence was received, UniProt was searched. The HP's properties have been recorded for investigation. Different parameters, which represent the properties of the HP as well as HP’s sequence in FASTA format, have been saved. In this HP of SC, the number of total amino acids is 534, designated as WN66 06778. Other parameters, such as locus, were assigned as KZV07215, likewise accession as KZV07215, and version numbers as KZV07215.1 (**Table 1**).

Table 1. Representation of distinctive attributes of HP derived from the NCBI protein database.

Properties	Hypothetical Protein
Definition	Hypothetical protein WN66_06778 (mitochondrion) [<i>Saccharomyces cerevisiae</i>]
Accession	KZV07215
Version	KZV07215.1
Amino acid	534
Organism	<i>Saccharomyces cerevisiae</i>
FASTA sequence	>KZV07215.1 hypothetical protein WN66_06778 (mitochondrion) [<i>Saccharomyces cerevisiae</i>] MVQRWLYSTNAKDIAVLVYFMLAIFSGMAGTAMSLIIRLELAAPGSQYLHGNSQLF NVLVVGHAVLMIFFLVMPALIGGFGNYLLPLMIGATDTAFPRINNIAFWVLPMLV CLVTSTLVESGAGTGWTVYPPLSSIQAHSGPSVDLAIFALHLTSISSLLGAINFIVTT LNMRTNGMTMHKLPLFVWSIFITAFLLLLSLPVLSAGITMLLLDRNFNTSFFEVAG GGDPILYEHLFWFFGHPEVYILIIPGFGIISHVVSTYSKKPVFGEISMVYAMASIGLL GFLVWSHHMYIVGLDADTRAYFTSATMIIAIPITGIKIFSWLATVYGGSSIRLATPMLY AIAFLFLFTMGGLTGVALANASLDVAFHDTYYVVGHFHYVLSMGAIFSLFAGYYY WSPQILGLNYNEKLAQIQFWLIFIGANVIFFPMHFLGINGMPRRIPDYPDAFAGWN YVASIGSFIATLSLFLFIYILDQLVNLNKNVNNKSVIYAKAPDFVESNTIFNLNTVK SSSIEFLLTSPPAVHSFNTPAVQS

NCBI (National Center for Biotechnology Information); HP (hypothetical protein).

3.2. Physicochemical Properties Characterization

The physicochemical characteristics of the selected hypothetical protein (HP) were evaluated using the ExPASy ProtParam server, and the results are summarized in **Table 2**. The HP consists of 534 amino acids with an estimated molecular weight of 63,295.44 Da. The theoretical isoelectric point (pI) was calculated to be 6.55, indicating that the protein is moderately acidic while retaining a substantial net positive charge at physiological pH. Protein stability is a critical determinant of its structural integrity and functional performance across biological systems. The instability index (II), a commonly used indicator of in vivo stability, predicts a protein to be stable when the score is below 40 [57]. The aliphatic index (AI), which reflects the relative volume of aliphatic side chains and correlates with thermal stability, was calculated as 114.51 showing the target protein's increased thermal stability [58]. The hydropathy value of every amino acid regarding the desired sequence is calculated, and after dividing this value by the total number, we can finally obtain the GRAVY value for the queued protein. When the value of GRAVY is negative, the protein is considered hydrophilic. But in terms of a hydrophobic protein, this value is found as positive. HP has a calculated value of 0.76, indicating that it is a hydrophobic protein. The extinction coefficient, according to the Beer-Lambert rule, measures as a proportionality constant that estimates the power of specific light that is captured by a protein [59]. It was therefore determined that the HP's extinction coefficient was 92250. The high extinction coefficient means that there is an abundance of tyrosine, tryptophan, and cysteine in the surroundings [57]. However, **Table 2** represents some other properties of the HP's physicochemical features.

Table 2. Representation of physicochemical characteristics of HP obtained from the ProtParam tool.

Parameters	Values
Total Number of Negatively Charged Residues (Asp + Glu)	22
Total Number of Positively Charged Residues (Arg + Lys)	19
Formula	C ₂₇₇₉ H ₄₁₉₄ N ₆₄₆ O ₇₁₂ S ₂₂
Total Number of Atoms	8353
Estimated Half-Life	30 hours (mammalian reticulocytes, in vitro) >20 hours (yeast, in vivo) >10 hours (<i>Escherichia coli</i> , in vivo)

HP (hypothetical protein), pI (isoelectric point).

3.3. Annotation of Functional Domain

Multiple servers confirmed that the HP contains a well-established conserved domain named Cytochrome c oxidase-like subunit I (COX-I) (**Supplementary Table 1**). Through asymmetric proton absorption and proton pumping, electrons from COX transfer to oxygen molecules to absorb energy from the cell's membrane potential. In *SC*, the complex is made up of 11 subunits. All respiring species share an elevated level of preservation in COX-I, one of the main subunits that is encoded within mitochondria. The fundamental protein COX-I is inserted with the redox cofactors copper and heme. Since COX enters the procedure of the assembly process from both sides of the membrane that covers the mitochondria, a high number of assembly components are needed for the enzyme to be assembled. Among the most precisely defined and most preserved assembly-related indicators, Shy1 (SURF-I homologue of yeast) is lost in *SC*. This causes a developmental deficit in non-

biodegradable forms of carbon and a substantial decline in cytochrome oxidase [60]. The hypothesis that Shy1 participates in the Heme A3 insertion phase is supported by recent research. Translational activators for SC COX-I mRNA include Pet309 and the mitochondrial splicing suppressor protein 51 (Mss51). Pet309, the COX-I mRNA activator, belongs to the protein family known as pentatricopeptide repeats (PPR), which is frequently linked to RNA metabolism [61]. The versatile pioneer protein Mss51, the second COX-I-specific activator, lacks any functional domains or patterns in its sequence. Several computational techniques have proposed that HP's unique domain serves as subunit I of cytochrome c oxidase.

3.4. Multiple Sequence Alignment and Analysis of Phylogeny

For conducting MSA (Multiple Sequence Alignment), the UniProt database acted like a BLAST server, returning the values of HP for proteins discovered. This tool was used to search the protein database for microorganisms having the greatest proportion of sequence similarity, the protein and organism name, and the length of amino acid sequences (**Supplementary Table 2**). Following that, the MAFFT tool was applied to perform sequence alignment. The phylogeny analysis was performed so that it could identify the evolutionary relationship of KZV07215.1 and its origin. The phylogenetic analysis was performed for the first 24 reviewed homologous protein sequences from different organisms. It was observed that the sequences of the HP and *Saccharomyces* COX-I are 99.3% similar, putting them in a similar clade of the neighbor-joining tree (**Figure 1**). Nevertheless, the tree also showed us that the COX-I protein of *Saccharomyces paradoxus*, *Candida glabrata*, and *Kluyveromyces lactis* has maximum sequence similarity with that of HP, 99.1%, 91.4%, and 91.4%, respectively. Moreover, *Wickerhamomyces canadensis* and *Vanderwaltozyma polyspora* COX-I proteins showed more than 80% sequence similarity with HP of SC (83.9% and 86.3%, respectively).

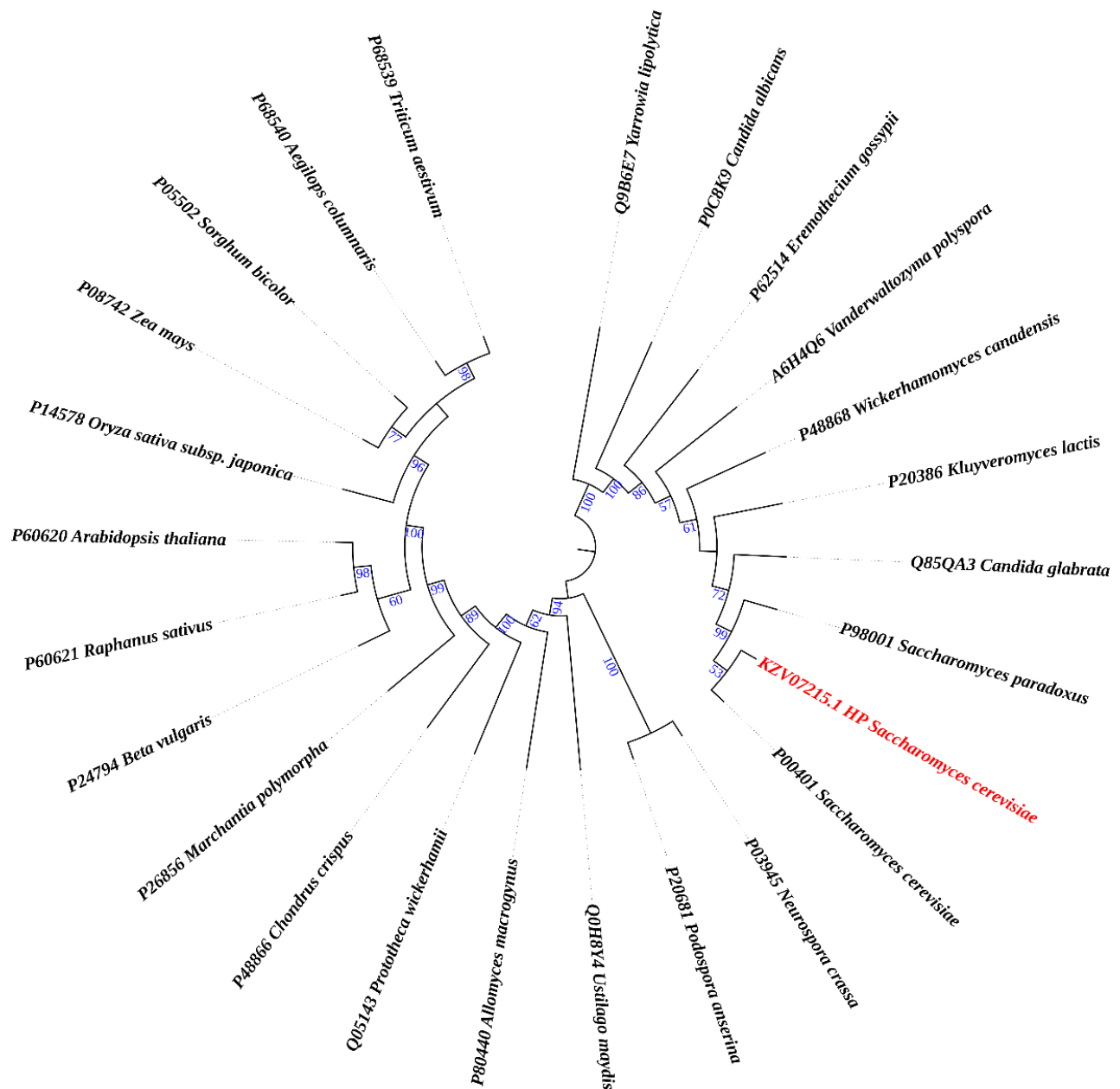


Figure 1. Phylogenetic tree revealing the ancestral relationship of the queued protein (marked as red) with other subunit I of COX proteins.

3.5. Secondary Structure Prediction

The 2D structure prediction was anticipated using a couple of techniques, such as PSIPRED and the SOPMA website. To summarize, the PSIPRED server revealed that our predicted protein structure will have its largest alpha-helix area, followed by random coils, and lastly extended strands (**Figure 2**). As predicted by PSIPRED, the SOPMA server concurred with it, stating that this HP might have contained a higher fraction of alpha-helix than prolonged strand or random coils (**Supplementary Table 3**).

3.6. Tertiary Structure Prediction

The I-TASSER platform was used to precisely anticipate HP models. I-TASSER generated a model for our target protein, selecting one with a TM-score of 0.98 ± 0.05 and a calculated RMSD of 3.6 ± 2.5 from all models produced (**Figure 3**). A TM value exceeding 0.5 indicates a structure with suitable topology, whereas a TM value below 0.17 suggests random similarity. Consequently, the YASARA server has increased the energy of the protein structure, while the GalaxyRefine server has enhanced the accuracy of its projected models. The program PyMOL was used to analyze and display the three-dimensional structures of the proposed and modified models. The YASARA energy-minimization server reduced the protein structural energy from -219693.9 to -247734.9 kJ/mol. The initial value was -2.59; however, after optimizing the energy of the HP, the final value was -0.48, indicating a comparatively more stable structure.

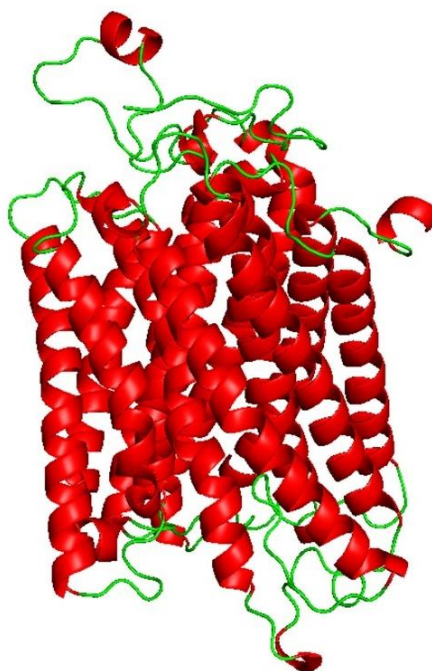


Figure 3. Predicted 3D configuration of the hypothetical protein via I-TASSER server after minimizing energy using YASARA and GalaxyRefine (visualized by PyMOL). The red-spiral represents alpha helices and no beta-sheet arrangements, while the horizontal ribbon stands for the HP's coil structure, in that order. HP stands for hypothetical protein.

3.7. Assessment of Structure Quality

According to the SAVES PROCHECK, 95.5% of the residues of amino acids were situated among a Ramachandran's top choice areas, whereas 3.7% were positioned in additional authorized regions (**Figure 4**). The ERRAT score of the I-TASSER predicted model is 92.381. According to the SWISS-MODEL, MolProbity gives I-TASSER's HP a value of 1.52, Ramachandran favored region of 97.18%, a QMEAN value of -3.36, and a QMEANDisCo Global of 0.77 ± 0.05 (**Supplementary Figure 1**). We examined the structure and decided on the I-TASSER model for further investigation.

According to ProSA, the predicted Z-score of the HP model is given a negative figure by I-TASSER: -5.71 (**Supplementary Figure 2**).

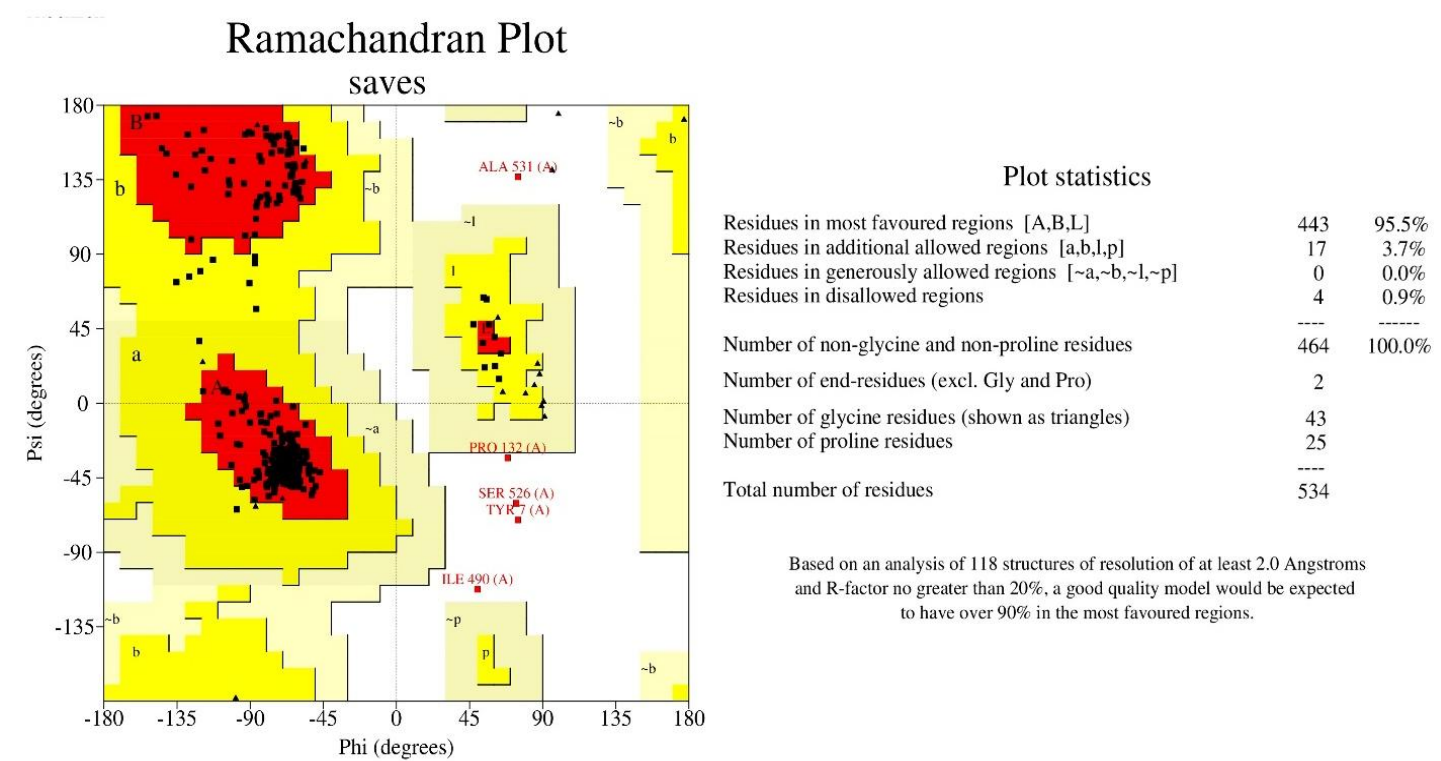


Figure 4. The model displayed through the I-TASSER server's Ramachandran plot. The beta-sheet area and other tertiary structural elements of the HP are represented by the first quadrant, where the alpha-helix area of the left and right hands is represented by the second and third quadrants, respectively. Furthermore, the residues in the most preferred, additional allowed, generously allowed, and prohibited regions are represented by the color regions in red, yellow, gray, and white, respectively.

3.8. Prediction of Subcellular Localization

Several servers, including Euk-mPLoc 2.0, TargetP-2.0, and MitPred, have predicted where the HP will be located within the cell. Various biological activities are connected to a variety of cellular sites [62]. Identifying HP’s location within the cell may shed light on the protein’s probable function. Also, this information could be used to develop a medicine that suppresses the action of the protein being investigated [62]. The researchers proposed that HP belongs to the mitochondrial protein (**Supplementary Table 4**).

3.9. Prediction of Binding Site

The model's binding site was evaluated using the FTSite server. The amino acid residues inside the binding area were also considered. The results were then shown using the program PyMOL (**Figure 5**). Characterizing the residues of the binding site is crucial for understanding the functionality of such sites. The active residues of the three main active areas of the model protein are identified using the FTSite estimation (**Supplementary Table 5**).

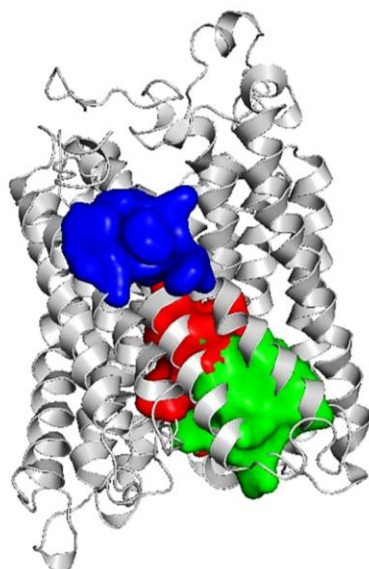


Figure 5. Representation of activation sites along with active amino acid molecules predicted by the FTase server. The protein is shown by the cyan hue; on the other hand, the red, green, and blue colors indicate the active sites I, II, and III, respectively.

Upon evaluating the active sites of the target protein using the FTSite server, around thirty amino acid residues were identified. Meanwhile, its active site had five functional amino acid residues, as identified by InterProScan: His-62, His-378, Ser-382, Ile-424, and Ala-461. Based on predictions from InterProScan, NCBI-CD Search, and Pfam, these residues are situated within the COX-I domains of the potential target protein. The HDock server was used to analyze the docking patterns of the ligand and the target protein. The target protein (WN66_06778) was conjugated with the ligand Heme A (HEA). A greater negative docking score suggests a more plausible binding model. Our investigation indicates that HP-HEA III has the greatest potential to establish a binding model, with a maximum docking score of -159.90. The probability of the two molecules joining is maximized when the confidence level exceeds 0.7. The likelihood of the two molecules joining is moderate to high when the confidence level ranges from 0.5 to 0.7 and minimal when the value falls below 0.5. The confidence score of HP-HEA III in this instance is 0.5493, indicating a substantial probability of molecular binding. HP-HEA II has a confidence score of roughly 0.5, indicating its propensity to bind. Generally, a closer alignment of the docking posture with the ligand's binding mode results in a reduced root mean square deviation (RMSD) seen throughout the experiment.

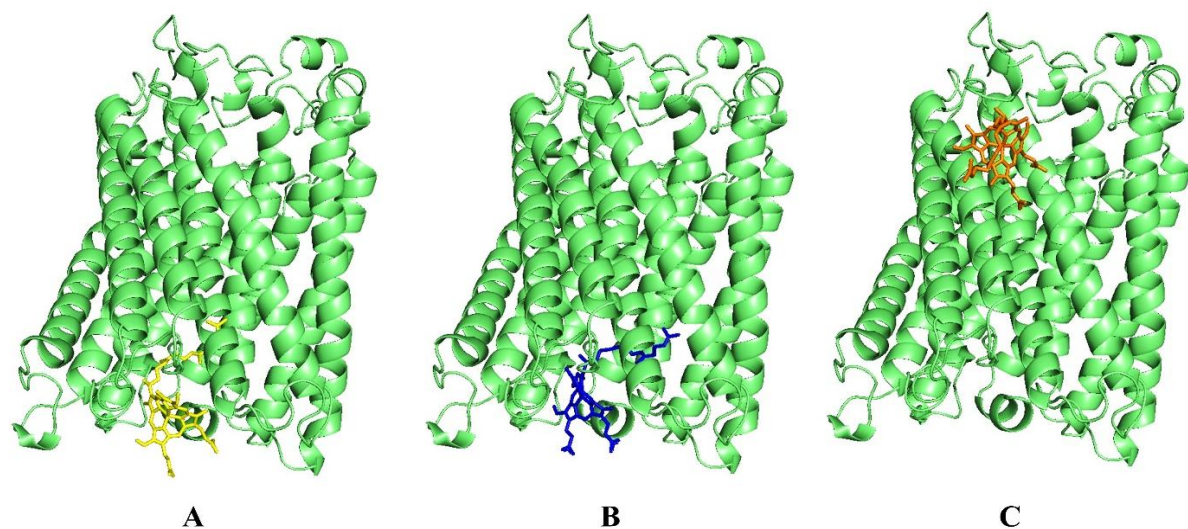


Figure 6. Molecular docking investigation of HP with Heme A. The site-specific docking was shown by the figure (A, B and C). Here, the ribbon that is displayed represents the HP and the strands depict the ligand (green color).

In our analysis, HP-HEA III has the lowest ligand RMSD value of 103.05 Å. The ligand and protein exhibited a substantial binding affinity. The analysis revealed a similarity between the HP and several interaction residues in the active site. The result aligns with the FTSite server's assessment of the active site. Results of the docking are illustrated in **Figure 6** and **Table 3**.

Table 3. Investigation of molecular docking of the HP with the Heme A.

Site ID	Interacting residues	Docking score	Confidence score	Ligand RMSD (Å)
HP-HEA I	HIS62, ALA63, MET66, ILE67, VAL71, TRP127, LEU247, VAL373, VAL374, PHE377, HIS378, LEU381, SER382, ARG438	-108.99	0.3057	127.63
HP-HEA II	SER33, ILE36, ARG37, LEU40, PHE55, VAL59, HIS62, TYR371, VAL374, ARG438, ARG439, ILE440	-147.16	0.4858	127.43
HP-HEA III	LEU17, MET20, LEU21, PHE24, PHE103, TRP104, PRO107	-159.90	0.5493	103.05

4. DISCUSSION

In microbial and eukaryotic genomes, hypothetical proteins (HPs) are a significant although unexplored component. Bioinformatic characterization is crucial for identifying preliminary functions and directing future empirical studies because many of them lack experimental validation at the protein level, despite their presence being obvious at the sequence level [63]. Understanding the functions of HPs is essential to fully comprehending genomic and proteomic landscapes since these proteins often represent novel structures, undiscovered regulatory mechanisms, or uncharacterized enzymatic activity [21, 26]. Additionally, their characterization can shed light on metabolic pathways, mechanisms linked to disease, and possible targets for treatment [64].

In this study, we performed a comprehensive in silico evaluation of the hypothetical mitochondrial protein WN66_06778 from *Saccharomyces cerevisiae*. Physicochemical analysis revealed a stable, moderately acidic, hydrophobic protein with a calculated molecular weight of 58.75 kDa, a theoretical pI of 6.55, and a GRAVY score of 0.760. Multiple localization servers consistently predicted its mitochondrial residency, aligning with its hydrophobic nature and suggested transmembrane features. Secondary structure predictions indicated that the protein is predominantly alpha-helical, consistent with known structures of mitochondrial membrane proteins, particularly members of the electron transport chain [65, 66]. Functional annotation through InterProScan, Pfam, and NCBI-CDD strongly identified conserved motifs corresponding to cytochrome c oxidase subunit I (COX-I). This prediction was reinforced by BLAST-based homology analysis, which revealed up to 99% sequence identity with known COX-I proteins from closely related yeast species (**Supplementary Table 2**). COX-I is a core catalytic subunit of Complex IV, integral for proton pumping and electron transfer to molecular oxygen during oxidative phosphorylation. Its biogenesis is tightly coordinated through mitochondrial and nuclear gene expression, translational regulation, and cofactor incorporation [67-70].

Mitochondrial COX assembly requires an intricate interplay of nuclear-encoded factors, including Pet309, Mss51, and Shy1, which orchestrate COX-I translation, heme A insertion, and holoenzyme formation [71-76]. Disruption of these processes, often studied in petite mutants, impairs respiratory capacity and highlights the central role of COX-I as the structural scaffold for assembly of the full complex. Previous studies confirm that when COX subunits or cofactors are unavailable, COX-I synthesis is downregulated to prevent accumulation of assembly intermediates [73, 74]. The protein examined in this study demonstrated strong alignment with these conserved features, supporting its predicted classification as a COX-I-like protein. Structural predictions generated using I-TASSER produced a high-confidence 3D model, which passed multiple validation criteria, including PROCHECK, QMEAN, ERRAT, and energy minimization via YASARA. The high TM-score (0.98 ± 0.05) indicates strong structural homology to experimentally resolved COX-I proteins. Active site mapping with FTSite identified approximately thirty putative functional residues, many of which aligned with known catalytic or ligand-binding positions within COX-I. Molecular docking with heme A, the primary cofactor for COX-I further substantiated this functional prediction. Among the docking complexes, HP-HEA III exhibited the strongest affinity (-159.90 kcal/mol) with superior confidence and lower RMSD values, indicating a stable and plausible interaction. Many interacting residues overlapped with the predicted active-site regions, reinforcing the protein's likely catalytic capacity rather than a purely structural or accessory role [77-80].

The functional relevance of COX-I is underscored by its association with numerous mitochondrial pathologies. Defects in COX-I assembly or function can lead to impaired oxidative phosphorylation, elevated reactive oxygen species production, and severe clinical manifestations such as mitochondrial encephalomyopathies, Leigh syndrome, and neonatal-onset COX deficiency syndromes [81]. Several human orthologues of yeast COX assembly factors, including SCO1 and SCO2, are known contributors to such disorders, demonstrating evolutionary conservation and biomedical significance [82-84]. Thus, the accurate annotation of mitochondrial HPs with COX-I-like features may have broader implications for understanding mitochondrial dysfunction across species.

Overall, our findings offer strong computational evidence that the hypothetical protein WN66_06778 is a COX-I like mitochondrial protein with plausible catalytic activity and functional relevance within the electron transport chain. This study underscores the utility of integrative bioinformatics in clarifying the roles of uncharacterized proteins and sets the stage for targeted experimental validation. Beyond its biological importance in *S. cerevisiae*, this protein may hold translational relevance for mitochondrial research, industrial biotechnology, and the study of oxidative phosphorylation disorders.

5. CONCLUSION

This study provides a comprehensive computational characterization of the hypothetical mitochondrial protein WN66_06778 from *Saccharomyces cerevisiae*. Through integrated physicochemical analysis, structural modeling, domain annotation, evolutionary profiling, and molecular docking, we identified strong evidence that this uncharacterized protein represents a cytochrome c oxidase subunit I (COX-I) like protein. The predicted structural stability, predominant α -helical architecture, mitochondrial localization, and high sequence homology with known COX-I proteins collectively support this functional classification. High-confidence 3D modeling and validation further reinforced the reliability of the predicted structure, while binding site analysis and docking with heme A demonstrated robust ligand interactions consistent with a catalytic role in mitochondrial Complex IV. These findings not only clarify the probable biological role of WN66_06778 but also underscore its importance in oxidative phosphorylation and mitochondrial energy metabolism. Given the central role of COX-I in electron transport and the association of its dysfunction with diverse mitochondrial disorders, the detailed annotation of this protein creates a foundation for future experimental validation. Overall, this work highlights the power of in silico approaches for functional prediction of hypothetical proteins and provides valuable insights that may support future studies in yeast bioenergetics, mitochondrial biology, and industrial or biomedical applications.

ACKNOWLEDGEMENTS

The authors would like to thank the Faculty of Biotechnology and Genetic Engineering, Sylhet Agricultural University, Sylhet-3100, Bangladesh, and the Department of Biochemistry and Molecular Biology, Gono Bishwabidyalay (University), Dhaka-1344, Bangladesh, for the technical support of this research.

ETHICS APPROVAL

Not applicable.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

FUNDING INFORMATION

This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors.

ARTIFICIAL INTELLIGENCE (AI) USE DISCLOSURE

The authors declare that artificial intelligence (AI) tools were used to assist in the preparation of this manuscript (approximately 10%) for improving language, grammar, and formatting. All content was reviewed and verified by the authors to ensure accuracy and compliance with ethical standards. The authors shall be solely responsible for any misconduct related to this work.

SUPPLEMENTARY MATERIALS

All supplementary materials referenced in this article are available in the online version.

REFERENCES

1. Parapouli M, Vasileiadis A, Afendra AS, Hatziloukas E. *Saccharomyces cerevisiae* and its industrial applications. *AIMS Microbiol.* 2020;6(1):1. doi:10.3934/microbiol.2020001
2. Ephrussi B. Action de l'acri flavine sur les levures. I. La mutation "petite colonie." *Ann Inst Pasteur.* 1949;76:351-357.
3. Tzagoloff A, Akai A, Needleman RB, Zulch G. Assembly of the mitochondrial membrane system: Cytoplasmic mutants of *Saccharomyces cerevisiae* with lesions in enzymes of the respiratory chain and in the mitochondrial ATPase. *J Biol Chem.* 1975;250(20):8236-8242.
4. Foury F, Roganti T, Lecrenier N, Purnelle B. The complete sequence of the mitochondrial genome of *Saccharomyces cerevisiae*. *FEBS Lett.* 1998;440(3):325-331. doi:10.1016/S0014-5793(98)01467-7
5. Bos JL, Heyting C, Borst P, Arnberg AC, van Bruggen EFJ. An insert in the single gene for the large ribosomal RNA in yeast mitochondrial DNA. *Nature.* 1978;275(5678):336-338. doi:10.1038/275336a0
6. Hensgens LA, Grivell LA, Borst P, Bos JL. Nucleotide sequence of the mitochondrial structural gene for subunit 9 of yeast ATPase complex. *Proc Natl Acad Sci U S A.* 1979;76(4):1663-1667. doi:10.1073/pnas.76.4.1663
7. Macino G, Tzagoloff A. Assembly of the mitochondrial membrane system: Partial sequence of a mitochondrial ATPase gene in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A.* 1979;76(1):131-135. doi:10.1073/pnas.76.1.131
8. Goffeau A, Barrell BG, Bussey H, et al. Life with 6000 genes. *Science.* 1996;274(5287):546-567. doi:10.1126/science.274.5287.546
9. Oliver SG, van der Aart QJM, Agostoni-Carbone ML, et al. The complete DNA sequence of yeast chromosome III. *Nature.* 1992;357(6373):38-46. doi:10.1038/357038a0
10. Oliver SG. From DNA sequence to biological function. *Nature.* 1996;379(6566):597-600. doi:10.1038/379597a0
11. Mewes HW, Albermann K, Bähr M, et al. Overview of the yeast genome. *Nature.* 1997;387(6632):7-8. doi:10.1038/387007a0
12. Hieter P, Boguski M. Functional genomics: It's all how you read it. *Science.* 1997;278(5338):601-602. doi:10.1126/science.278.5338.601

13. Hasan MR, Tithi SR, Khatun B, Chowdhury UR, Mahmud A. In silico characterization of hypothetical protein AAFM48_25015 from *Burkholderia pseudomallei*: Structural and functional analysis. *J Biosci Public Health*. 2025;1(3):28-44. doi:10.5455/JBPH.2025.13
14. Babcock GT, Wikström M. Oxygen activation and the conservation of energy in cell respiration. *Nature*. 1992;356(6367):301-309. doi:10.1038/356301a0
15. Guo Q, Li J, Wang MR, et al. Multidimensional engineering of *Saccharomyces cerevisiae* for efficient heme production. *Metab Eng*. 2024;85:46-60. doi:10.1016/j.ymben.2024.01.004
16. Capaldi RA. Structure and function of cytochrome c oxidase. *Annu Rev Biochem*. 1990;59:569-596. doi:10.1146/annurev.bi.59.070190.003033
17. Brunori M, Antonini G, Malatesta F, Sarti P, Wilson MT. Cytochrome-c oxidase: Subunit structure and proton pumping. *Eur J Biochem*. 1987;169(1):1-8. doi:10.1111/j.1432-1033.1987.tb13554.x
18. Saraste M. Structural features of cytochrome oxidase. *Q Rev Biophys*. 1990;23(4):331-366. doi:10.1017/S0033583500005648
19. Maréchal A, Meunier B, Lee D, Orengo C, Rich PR. Yeast cytochrome c oxidase as a model system. *Biochim Biophys Acta Bioenerg*. 2012;1817(4):620-628. doi:10.1016/j.bbabi.2011.09.009
20. Sanchez R, Sali A. Large-scale protein structure modeling of the *Saccharomyces cerevisiae* genome. *Proc Natl Acad Sci U S A*. 1998;95(23):13597-13602. doi:10.1073/pnas.95.23.13597
21. Barua H, Hasan MR, Mardiyah RT, et al. Developing a novel multi-epitope subunit vaccine against monkeypox virus through immunoinformatics. *Vacunas*. 2025;500490. doi:10.1016/j.vacun.2025.500490
22. Imai K, Asakawa N, Tsuji T, et al. SOSUI-GramN: High-performance prediction for subcellular localization of Gram-negative proteins. *Bioinformatics*. 2008;2(9):417-421.
23. Marchler-Bauer A, Bo Y, Han L, et al. CDD/SPARCLE: Functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res*. 2017;45(D1):D200-D203. doi:10.1093/nar/gkw1129
24. Quevillon E, Silventoinen V, Pillai S, et al. InterProScan: Protein domains identifier. *Nucleic Acids Res*. 2005;33(suppl 2):W116-W120. doi:10.1093/nar/gki442
25. Gough J, Karplus K, Hughey R, Chothia C. Assignment of homology to genome sequences using HMM libraries. *J Mol Biol*. 2001;313(4):903-919. doi:10.1006/jmbi.2001.5080
26. Paul M, Mashrur N, Roy A, Hasan R. Computational characterization of *Mycobacterium tuberculosis* hypothetical protein as L,D-transpeptidase. *J Bioinform Res Rev*. 2025;2(3):168-185. doi:10.5455/JBRR.20250710062830
27. Morgat A, Lombardot T, Coudert E, et al. Enzyme annotation in UniProtKB using Rhea. *Bioinformatics*. 2020;36(6):1896-1901. doi:10.1093/bioinformatics/btz817
28. Pundir S, Martin MJ, O'Donovan C; UniProt Consortium. UniProt tools. *Curr Protoc Bioinformatics*. 2016;53:1-29.
29. Rozewicki J, Li S, Amada KM, Standley DM, Katoh K. MAFFT-DASH: Integrated protein sequence-structure alignment. *Nucleic Acids Res*. 2019;47(W1):W5-W10. doi:10.1093/nar/gkz342
30. Guindon S, Dufayard JF, Lefort V, et al. PhyML 3.0: New algorithms for maximum-likelihood phylogenies. Accessed May 14, 2025. <http://www.lirmm.fr/~gascuel>
31. Letunic I, Bork P. iTOL v5: Online tool for phylogenetic tree display. *Nucleic Acids Res*. 2021;49(W1):W293-W296. doi:10.1093/nar/gkab301
32. Combet C, Blanchet C, Geourjon C, Deleage G. NPS@: Network protein sequence analysis. *Trends Biochem Sci*. 2000;25(3):147-150.
33. Geourjon C, Deleage G. SOPMA: Improved secondary structure prediction. *Bioinformatics*. 1995;11(6):681-684. doi:10.1093/bioinformatics/11.6.681
34. Buchan DWA, Jones DT. PSIPRED protein analysis workbench: 20-year update. *Nucleic Acids Res*. 2019;47(W1):W402-W407.
35. Jones DT. Protein secondary structure prediction with PSSM. *J Mol Biol*. 1999;292(2):195-202. doi:10.1006/jmbi.1999.3091
36. Yang J, Zhang Y. I-TASSER server: Developments in protein prediction. *Nucleic Acids Res*. 2015;43(W1):W174-W181. doi:10.1093/nar/gkv342
37. Zhang C, Freddolino PL, Zhang Y. COFACTOR: Improved protein function prediction. *Nucleic Acids Res*. 2017;45(W1):W291-W299.
38. Zheng W, Zhang C, Li Y, Pearce R, Bell EW, Zhang Y. Folding nonhomologous proteins via deep learning contact maps with I-TASSER simulations. *Cell Reports Methods*. 2021;1(3):100014.
39. Ahmed N, Rani NA, Robin TB, et al. Multi-epitope subunit vaccine design against *Toxoplasma gondii*. *Mol Biochem Parasitol*. 2024;260:111655. doi:10.1016/j.molbiopara.2024.111655
40. Krieger E, Joo K, Lee J, et al. Improving accuracy in homology modeling: CASP8 methods. *Proteins*. 2009;77(S9):114-122. doi:10.1002/prot.22570
41. Mashrur MN, Hoque MN, Hasnat S, et al. Multi-epitope vaccine candidate for Marburg virus. *Vacunas*. 2025;26(3). doi:10.1016/j.vacun.2025.500461
42. DeLano WL. PyMOL: An open-source molecular graphics tool. *CCP4 Newsl Protein Crystallogr*. 2002;40:82-92.

43. Yuan S, Chan HCS, Hu Z. Using PyMOL for computational drug design. *WIREs Comput Mol Sci*. 2017;7(2):e1298. doi:10.1002/wcms.1298
44. Waterhouse A, Bertoni M, Bienert S, et al. SWISS-MODEL: Homology modeling of protein structures and complexes. *Nucleic Acids Res*. 2018;46(W1):W296-W303. doi:10.1093/nar/gky427
45. Laskowski RA, Rullmann JAC, MacArthur MW, Kaptein R, Thornton JM. AQUA and PROCHECK-NMR: Structure quality checks. *J Biomol NMR*. 1996;8(4):477-486. doi:10.1007/BF00228148
46. Studer G, Rempfer C, Waterhouse AM, Gumienny R, Haas J, Schwede T. QMEANDisCo: Distance constraints for model quality estimation. *Bioinformatics*. 2020;36(6):1765-1771
47. Benkert P, Künzli M, Schwede T. QMEAN server for protein model quality estimation. *Nucleic Acids Res*. 2009;37(suppl 2):W510-W514. doi:10.1093/nar/gkp322
48. Dönnes P, Höglund A. Predicting protein subcellular localization. *Genomics Proteomics Bioinformatics*. 2004;2(4):209-215.
49. Chou KC, Shen HB. Euk-mPLoc 2.0: Subcellular localization prediction. *PLoS One*. 2010;5(4):e9931. doi:10.1371/journal.pone.0009931
50. Armenteros JJA, Salvatore M, Emanuelsson O, et al. Targeting peptides via deep learning. *Life Sci Alliance*. 2019;2(5):e201900429. doi:10.26508/lsa.201900429
51. Garg A, Bhasin M, Raghava GPS. SVM-based subcellular localization of human proteins. *J Biol Chem*. 2005;280(15):14427-14432. doi:10.1074/jbc.M413731200
52. Ngan CH, Hall DR, Zerbe B, Grove LE, Kozakov D, Vajda S. FTSite: Ligand binding site detection. *Bioinformatics*. 2012;28(2):286-287. doi:10.1093/bioinformatics/btr651
53. Tian W, Chen C, Lei X, Zhao J, Liang J. CASTp 3.0: Atlas of protein surface topography. *Nucleic Acids Res*. 2018;46(W1):W363-W367. doi:10.1093/nar/gky473
54. Khamis MA, Gomaa W, Ahmed WF. Machine learning in computational docking. *Artif Intell Med*. 2015;63(3):135-152. doi:10.1016/j.artmed.2015.02.004
55. Khamis MA, Galal B, Gomaa W. Deep learning competing with random forest in docking. In: *2023 11th International Japan-Africa Conference on Electronics, Communications, and Computations*. IEEE; 2023:62-67.
56. Berman HM, Westbrook J, Feng Z, et al. The Protein Data Bank. *Nucleic Acids Res*. 2000;28(1):235-242. doi:10.1093/nar/28.1.235
57. Gasteiger E, Hoogland C, Gattiker A, et al. Protein identification tools on ExPASy. In: *Proteomics Protocols Handbook*. 2005:571-607. doi:10.1385/1-59259-890-0:571
58. Enany S. Structural and functional analysis of hypothetical proteins of *Clostridium tetani*. *J Infect Public Health*. 2014;7(4):296-307. doi:10.1016/j.jiph.2014.04.007
59. Herzog B, Schultheiss A, Giesinger J. Validity of Beer–Lambert law for sunscreens. *Photochem Photobiol*. 2018;94(2):384-389. doi:10.1111/php.12862
60. Mashkevich G, Repetto B, Glerum DM, Jin C, Tzagoloff A. SHY1 encodes a mitochondrial protein required for respiration. *J Biol Chem*. 1997;272(22):14356-14364. doi:10.1074/jbc.272.22.14356
61. Tavares-Carreón F, Camacho-Villasana Y, Zamudio-Ochoa A, Shingu-Vazquez M, Torres-Larios A, Perez-Martinez X. PPR motifs in Pet309 required for translation of COX1 mRNA. *J Biol Chem*. 2008;283(3):1472-1479. doi:10.1074/jbc.M708370200
62. Yu C, Chen Y, Lu C, Hwang J. Prediction of protein subcellular localization. *Proteins*. 2006;64(3):643-651. doi:10.1002/prot.21028
63. Lubec G, Afjehi-Sadat L, Yang JW, John JP. Searching for hypothetical proteins. *Prog Neurobiol*. 2005;77(1-2):90-127. doi:10.1016/j.pneurobio.2005.10.002
64. Sen T, Verma NK. Functional annotation of hypothetical proteins in *Shigella flexneri*. *Genes (Basel)*. 2020;11(3):340. doi:10.3390/genes11030340
65. Miah ML, Hossain MF, Uddin ME, et al. Proteomic insights into cigarette smoking health impacts. *Bioresearch Commun*. 2025;11(1):1651-1678. doi:10.3329/brc.v11i1.78881
66. Hossain H, Chowdhury MSR, Khan SS, et al. Emerging zoonotic diseases and One Health. *Pak Vet J*. 2025;45(1):36-47. doi:10.29261/pakvetj/2025.115
67. Lipinski KA, Kaniak-Golik A, Golik P. Maintenance and expression of the yeast mitochondrial genome. *Biochim Biophys Acta Bioenerg*. 2010;1797(6-7):1086-1098. doi:10.1016/j.bbabi.2010.02.029
68. Naithani S, Saracco SA, Butler CA, Fox TD. Interactions among COX1, COX2, COX3 translational activators. *Mol Biol Cell*. 2003;14(1):324-333. doi:10.1091/mbc.E02-06-0330
69. Neupert W, Herrmann JM. Translocation of proteins into mitochondria. *Annu Rev Biochem*. 2007;76:723-749. doi:10.1146/annurev.biochem.76.052705.163409
70. Dienhart MK, Stuart RA. Aac2 association with bc1-COX supercomplex and TIM23 machinery. *Mol Biol Cell*. 2008;19(9):3934-3943. doi:10.1091/mbc.e08-04-0382
71. Pel HJ, Tzagoloff A, Grivell LA. Identification of nuclear genes required for COX1 expression. *Curr Genet*. 1992;21:139-146.

72. Tzagoloff A, Dieckmann C. PET genes of *Saccharomyces cerevisiae*. *Microbiol Rev*. 1990;54(3):211-225. doi:10.1128/mr.54.3.211-225.1990
73. Khalimonchuk O, Bestwick M, Meunier B, Watts TC, Winge DR. Formation of redox cofactor centers during Cox1 maturation. *Mol Cell Biol*. 2010;30(4):1004-1017. doi:10.1128/MCB.01035-09
74. Khalimonchuk O, Bird A, Winge DR. Pro-oxidant intermediate in cytochrome oxidase assembly. *J Biol Chem*. 2007;282(24):17442-17449. doi:10.1074/jbc.M702588200
75. Mick DU, Fox TD, Rehling P. Cytochrome c oxidase assembly regulates mitochondrial translation. *Nat Rev Mol Cell Biol*. 2011;12(1):14-20. doi:10.1038/nrm3029
76. Fontanesi F, Soto IC, Barrientos A. Regulation of cytochrome c oxidase biogenesis. *IUBMB Life*. 2008;60(9):557-568. doi:10.1002/iub.93
77. Lodolo EJ, Kock JLF, Axcell BC, Brooks M. Yeast in beer brewing. *FEMS Yeast Res*. 2008;8(7):1018-1036. doi:10.1111/j.1567-1364.2008.00428.x
78. Buijs NA, Siewers V, Nielsen J. Advanced biofuel production by *S. cerevisiae*. *Curr Opin Chem Biol*. 2013;17(3):480-488. doi:10.1016/j.cbpa.2013.03.035
79. Kim JS, Kim JW, Shim W, Kim JW, Park KH, Pek UH. Flavor-enhancing yeast extract using high-RNA *S. cerevisiae* strain. *Korean J Food Sci Technol*. 1999;31(2):475-481.
80. Kolkman A, Slijper M, Heck AJR. Proteomics technologies in *S. cerevisiae*. *Trends Biotechnol*. 2005;23(12):598-604. doi:10.1016/j.tibtech.2005.09.001
81. Wallace DC, Fan W, Procaccio V. Mitochondrial energetics and therapeutics. *Annu Rev Pathol*. 2010;5:297-348. doi:10.1146/annurev.pathol.4.110807.092314
82. Barrientos A, Barros MH, Valnot I, Rötig A, Rustin P, Tzagoloff A. Cytochrome oxidase in health and disease. *Gene*. 2002;286(1):53-63. doi:10.1016/S0378-1119(01)00814-1
83. Khatun Z, Saikat S, Haque S. Src family kinases in epidermal homeostasis, wound healing, and tumorigenesis. *J Biosci Public Health*. 2025;1(2):27-40. doi:https://doi.org/10.5455/JBPH.2025.08
84. Leary SC, Cobine PA, Kaufman BA, et al. SCO1 and SCO2 roles in copper homeostasis. *Cell Metab*. 2007;5(1):9-20. doi:10.1016/j.cmet.2006.11.012