**RESEARCH ARTICLE**

# *In-silico* structural modelling of cytochrome complex proteins of white turmeric (*Curcuma zedoaria* (Christm.) Roscoe

**Chandra Sekhar Tripathy, Bhaskar Chandra Sahoo, Manaswini Dash, Debasis Sahoo, Suprava Sahoo** & **Basudeba Kar***

Centre for Biotechnology, Siksha 'O' Anusandhan Deemed to be University, Bhubaneswar 751 003, Odisha, India
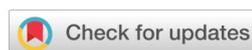
*Email: basu.cbt@gmail.com, supi.sos2000@gmail.com

**OPEN ACCESS**

## Abstract

*Curcuma zedoaria* (Christm.) Roscoe (white turmeric) is a perennial herbaceous plant of family Zingiberaceae and mainly found in the wild areas of tropical and subtropical regions worldwide. The cytochrome proteins in plants play important roles in promoting their growth and development, as well as protecting them from stresses and diseases. Cytochrome proteins like psbF, psbE, petB, petD, petN, petG and ccsA play important roles in degradation of misfolded proteins, ATP formation, cyclic electron flow and biogenesis of c-type cytochrome of *C. zedoaria*. However, due to lack of structural availability of these *C. zedoaria* cytochrome proteins in structural databases, the physiochemical parameters of sequences were estimated using Expasy ProtParam web tool. Self-Optimized Prediction Method with Alignment (SOPMA) server and MODELLER version 9.23 were used for modelling along with Qualitative Model Energy Analysis (QMEAN) and Protein Structure Analysis (ProSA) servers were implemented for validating the secondary and tertiary structures of these proteins. The obtained QMEAN4 values of the modelled cytochrome proteins were -2.04, -1.20, -3.01, -1.57, -2.11, -1.74 and -12.87. The Z-scores obtained from ProSA server were 0.5, -0.83, -1.5, -0.58, -0.02, 0.14 and -3.73. All seven modelled structures have been submitted to protein model database (PMDB). The derived results will be helpful in further investigations towards determining the crystal structure of the hypothetical proteins, structural motifs, physiochemical properties and also protein-protein interaction studies of various cytochrome proteins.

## Keywords

3D-structures, Ramachandran plot, Cytochrome proteins, Structural validation

## Introduction

*Curcuma zedoaria* (white turmeric) is one of the rhizomatous medicinal herbs that belongs to Zingiberaceae family. This plant is native to Bangladesh, India and Sri Lanka and is also cultivated in Brazil, China, Japan, Nepal and Thailand (1). *C. zedoaria* has been used in many folk medicines, especially in Ayurveda for the treatment of menstrual disorders, nausea, stomach diseases as well as carcinoma (2). The rhizome of the plant is used for its carminative, rubefacient, expectorant, diuretic, demulcent and stimulant properties while the root is used for the treatment of dyspepsia, flatulence, cough, cold and fever by the rural inhabitants (3). It has been also used in treating liver cancer, chronic pelvic inflammation, coronary heart disease, and helps to prevent leukopenia caused due to cancer therapies (4). In spite of having various medicinal importance, many works have not yet been done in this plant like

availability of cytochrome proteins structures. Cytochrome proteins play a very important role in regulating electron transport chain and redox catalysis in living organisms. In plants, cytochrome b559 subunit beta (psbF) and cytochrome b559 subunit alpha (psbE) proteins are the key components of photosystem II (PSII) reaction centre and participate in secondary electron transfer pathways as well as protect PSII against photoinhibition (5). Proteins such as cytochrome b6 (petB), cytochrome b6-f complex subunit 4 (petD), cytochrome b6-f complex subunit 8 (petN) and cytochrome b6-f complex subunit 5 (petG) are the intermediate cytochrome proteins which are transferred between photosystem II (PSII) and photosystem I (PSI) (6). The protein cytochrome C Biogenesis protein (ccsA) is required during biogenesis of c-type cytochromes (cytochrome c6 and cytochrome f) at the step of heme attachment. This information was available in Uniprot database (7). However, these cytochrome proteins of *C. zedoaria* are lacking 3D structures in protein databases.

Protein structures are very essential for the elucidation of various molecular mechanisms underlying the biological processes. Generally quality protein structures are developed by using X-ray crystallography and Nuclear Magnetic Resonance (NMR) spectroscopy which are very time consuming, costly and difficult processes specifically for membrane proteins (8). However, *in-silico* 3D structure prediction of proteins using homology modelling is an emerging alternative technique to develop protein structures (8). It is also essential to know the physiochemical properties for better understanding about the protein structures. Expasy ProtParam is one of the important tools for investigating the physiochemical properties (8, 9). Secondary structure prediction can be also achieved by the SOPMA server using the consensus multiple alignments. This method correctly predicts 69.5% of amino acids for a three-state description of the secondary structure (alpha-helix, beta-sheet and coil) (10). Further, MODELLER is well known for homology modelling and comparative tertiary structure prediction of proteins (11,12). For validation of the modelled structures, PROCHECK, Quantitative Model Energy Analysis (QMEAN) and Protein Structure Analysis (ProSA) servers are standard web servers for the validation of the modelled protein 3D structures (13,14).

In the present study, the 7 cytochrome proteins were targeted for *in-silico* structure prediction through various tools such as Expasy ProtParam, SOPMA, MODELLER, and RaptorX. Further, this study was also deigned to validate the 3D structures using WHAT IF, PROCHECK, QMEAN and ProSA servers.

## Materials and Methods

### Sequence availability

Unreviewed cytochrome protein sequences such as psbF, psbE, petB, petD, petN, petG and ccsA were retrieved in fasta format from UniProt database (7).

### Primary structure prediction

Primary structure prediction of proteins was carried out using the Expasy ProtParam web server (15). This was resulted with various physiochemical properties like theoretical isoelectric point (pI), molecular weight, total number of positive (+R) and negative residues (-R), extinction coefficient (Ec), instability index (II), aliphatic index (AI) and grand average of hydropathy (GRAVY).

### Secondary structure prediction

The secondary structures of the targeted proteins were predicted by SOPMA (10). The parameters of the server were set as follows for all the proteins: numbers of conformational states to 4 (helix, sheet, turn and coil), similarity threshold to 8 and window width to 17.

### Homology modelling

All the targeted protein sequences were taken as query for BLAST (Basic local alignment search tool) (16) using blastp tool. The protein sequences were used one by one as input data for the blastp tool against PDB database (17) to identify homologous structure. For homology modelling, MODELLER version 9.23 (11) was used. Further advance modelling and loop refinement were also carried out. As no blastp hits were found for cytochrome C Biogenesis protein, RaptorX **(18)** web server was employed for 3D structural modelling of protein. The 3D structures were visualised in BIOVIA Discovery studio 2020 tool.

### Model evaluation

The validation of 3D structures was done using online servers like WHAT IF web server to get fixed PDB structure and PROCHECK server (19) for Ramachandran plots. QMEAN server (20) was employed for composite scoring function of 3D structures to estimate both local (i.e., per residue) and global (i.e., for the entire structure) errors. The ProSA tool was used to evaluate overall quality of each model. The results of Z-score reflected the quality of modelled proteins in comparison to all other experimental protein structures which were estimated by X-ray crystallography and NMR (21). Quality protein structures were submitted to PMDB.

## Results

### Sequence retrieval

The UniProt database was used to retrieve seven unreviewed cytochrome protein sequences such as the psbF (39 bp), psbE (83 bp), petB (217 bp), petD (160 bp), petN (29 bp), petG (37 bp) and ccsA (329 bp). The retrieval protein sequence information was summarised in Table 1.

### Primary structure prediction

ProtParam tool of Expasy server was employed for primary structure prediction. This analysis resulted with physiochemical properties of the sequences. The molecular weight of the proteins ranged from 3169.83 to 37482.02 Daltons (Table 2). The pI values were from 4.60 to 10.74 for the cytochrome proteins. The GRAVY score was the lowest (0.076) for psbE. Additionally, petB, ccsA, petD, psbF, petG and petN were resulted with GRAVY score 0.503, 0.528, 0.605, 0.646, 1.114 and 1.545 respectively (Table 2). Other physiochemical properties like -R, +R, EC,

Table 1. Information of retrieved cytochrome proteins of *C. zedoaria*

| Sl. No. | Protein name | UniProt entry | Gene id | Length in base pairs |
|---|---|---|---|---|
| 1 | Cytochrome b559 subunit beta | A0A6H0ROB7 | psbF | 39 |
| 2 | Cytochrome b559 subunit alpha | A0A6H0R216 | psbE | 83 |
| 3 | Cytochrome b6 | A0A6H0QZU7 | petB | 217 |
| 4 | Cytochrome b6-f complex subunit 4 | A0A6H0QZF7 | petD | 160 |
| 5 | Cytochrome b6-f complex subunit 8 | A0A6H0QZA2 | petN | 29 |
| 6 | Cytochrome b6-f complex subunit 5 | A0A6H0R1P1 | petG | 37 |
| 7 | Cytochrome C Biogenesis protein | A0A6H0QZW8 | ccsA | 329 |

II, and AI of petB, petD, petN, petG and ccsA were shown in Table 2.

### Secondary structure prediction

Table 2. Physiochemical properties of predicted primary structures

| Sl. No. | Protein name | Length in base pairs | Mol. wt. | pI | -R | +R | Ec | II | AI | GRAVY |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Cytochrome b559 subunit beta | 39 | 4424.25 | 10.74 | 1 | 3 | 6990 | 26.50 | 107.44 | 0.646 |
| 2. | Cytochrome b559 subunit alpha | 83 | 9330.49 | 4.60 | 9 | 5 | 15470 | 42.32 | 84.58 | 0.076 |
| 3. | Cytochrome b6 | 217 | 24335.72 | 9.12 | 12 | 16 | 42400 | 35.20 | 103.23 | 0.503 |
| 4. | Cytochrome b6-f complex subunit 4 | 160 | 17520.93 | 6.56 | 11 | 11 | 22460 | 44.27 | 121.19 | 0.605 |
| 5. | Cytochrome b6-f complex subunit 8 | 29 | 3169.83 | 5.59 | 1 | 1 | 11000 | 17.38 | 141.03 | 1.545 |
| 6. | Cytochrome b6-f complex subunit 5 | 37 | 4170.02 | 4.56 | 3 | 2 | 2980 | 29.33 | 155.41 | 1.114 |
| 7. | Cytochrome C Biogenesis protein | 329 | 37482.02 | 9.29 | 12 | 21 | 77600 | 32.85 | 111.43 | 0.528 |

Mol. wt.- molecular weight (Daltons), pI- Isoelectric point, -R- Number of negative residues, +R- Number of positive residues, EC- Extinction coefficient at 280 nm, II- Instability Index, AI- Aliphatic Index, GRAVY- Grand average of hydropathy.

The secondary structure of predicted proteins was analysed by SOPMA server which was resulted with alpha helix, extended strand, beta turn and random coil. In this prediction of secondary structure, alpha helices were found as 12.82%, 31.33%, 62.67%, 29.38%, 68.97%, 29.73% and 48.94% in psbF, psbE, petB, petD, petN, petG and ccsA respectively (Table 3). It was followed by extended strands of 33.33%, 32.53%, 8.29%, 23.75%, 13.79%, 37.84% and 17.02% (Table 3). The beta turns were resulted with 5.13%, 7.23%, 5.53%, 3.75%, 10.34%, 8.11% and 3.95% (Table 3).

The random coils were found as 48.72%, 28.92%, 23.50%, 43.12%, 6.90%, 24.32% and 30.09% (Table 3). The cytochrome proteins revealed the cardinal nature of helix and coiling type reflects the more compact with strong attachment as well as existence in transmembrane region of the modelled proteins. The detailed secondary structure prediction information was listed in Table 3.

### Homology modelling and Structural validation

Here we have taken protein sequences in fasta format from Uniprot and searched using BLAST server under blastp tool one by one and obtained the following results. The similarity search of query sequences using blastp resulted with various similar proteins. The blastp search of psbF resulted with best similarity to 3JCU_F (query cover: 100%, identity: 94.87%), 5MDX_F (query cover: 100%, identity: 100.00%) and 5XNL_F (query cover: 100%, identity: 97.44%). The protein psbE resulted with best similarity to 3JCU_E (query cover: 100%, identity: 97.59%), 5XNL_E (query cover: 100%, identity: 95.18%) and 6KAC_E (query cover: 96%, identity: 73.75%). The protein petB resulted with best similarity to 1Q90_B (query cover: 99%, identity: 87.44%), 1VF5_A (query cover: 99%, identity: 81.86%) and 6RQF_A (query cover: 99%, identity: 95.35%). petD resulted with best similarity to 2E74_B (query cover: 100%, identity: 80.00%), 2ZT9_B (query cover: 100%, identity: 79.38%) and 6RQF_B (query cover: 100%, identity: 95.62 %). petN resulted with best similarity to 1VF5_H (query cover: 100%, identity: 63.33%), 2ZT9_H (query cover: 100%, identity: 58.62%) and 6R9F_H (query cover: 100%, identity: 100.00%). petG resulted with best similarity to 1Q90_G (query cover: 89%, identity: 75.76%), 2ZT9_G (query cover: 94%, identity:

Table 3. Secondary structure prediction summary of cytochrome proteins in *C. zedoaria*

| Secondary structure parameters | Cytochrome b559 subunit beta | Cytochrome b559 subunit alpha | Cytochrome b6 | Cytochrome b6-f complex subunit 4 | Cytochrome b6-f complex subunit 8 | Cytochrome b6-f complex subunit 5 | Cytochrome C Biogenesis protein |
|---|---|---|---|---|---|---|---|
| Alpha helix | 12.82% | 31.33% | 62.67% | 29.38% | 68.97% | 29.73% | 48.94% |
| $3_{10}$ helix | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| Pi helix | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| Beta Bridge | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |

| | | | | | | |
|---|---|---|---|---|---|---|
| Extended strand | 33.33% | 32.53% | 8.29% | 23.75% | 13.79% | 37.84% | 17.02% |
| Beta turn | 5.13% | 7.23% | 5.53% | 3.75% | 10.34% | 8.11% | 3.95% |
| Bend region | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| Random coil | 48.72% | 28.92% | 23.50% | 43.12% | 6.90% | 24.32% | 30.09% |
| Ambiguous state | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| Other states | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |

62.86%) and 6RQF_G (query cover: 100%, identity: 100.00%). However, no template sequences were found for ccsA using blastp tool.

Here, MODELLER 9.23 version used for building 3D structures of proteins using the blastp results. As no template sequence was found for ccsA protein, RaptorX online server has been used for modelling. The modelled 3D protein structures were shown in Fig. 1. For structural conformation of the 3D-structures, PROCHECK server was used for Ramachandran plot analysis as it provides a conven-
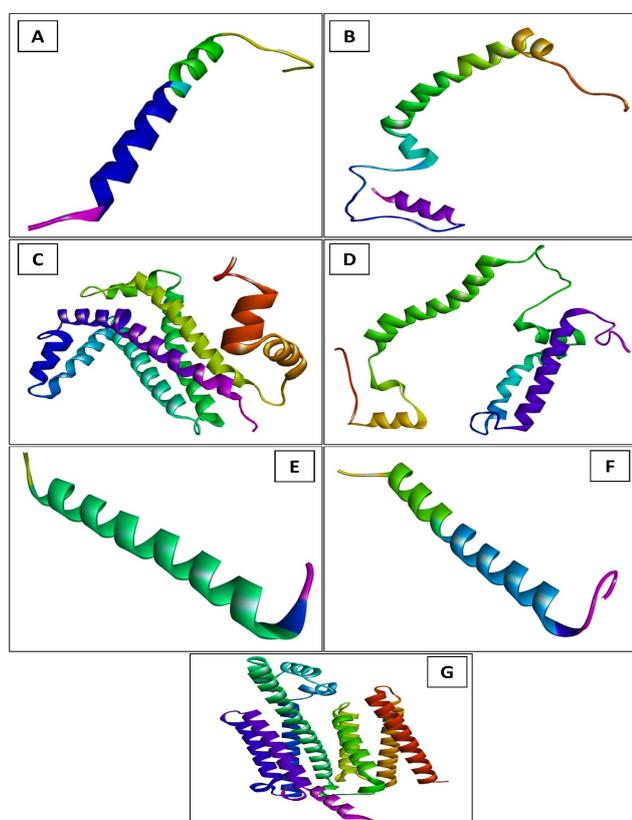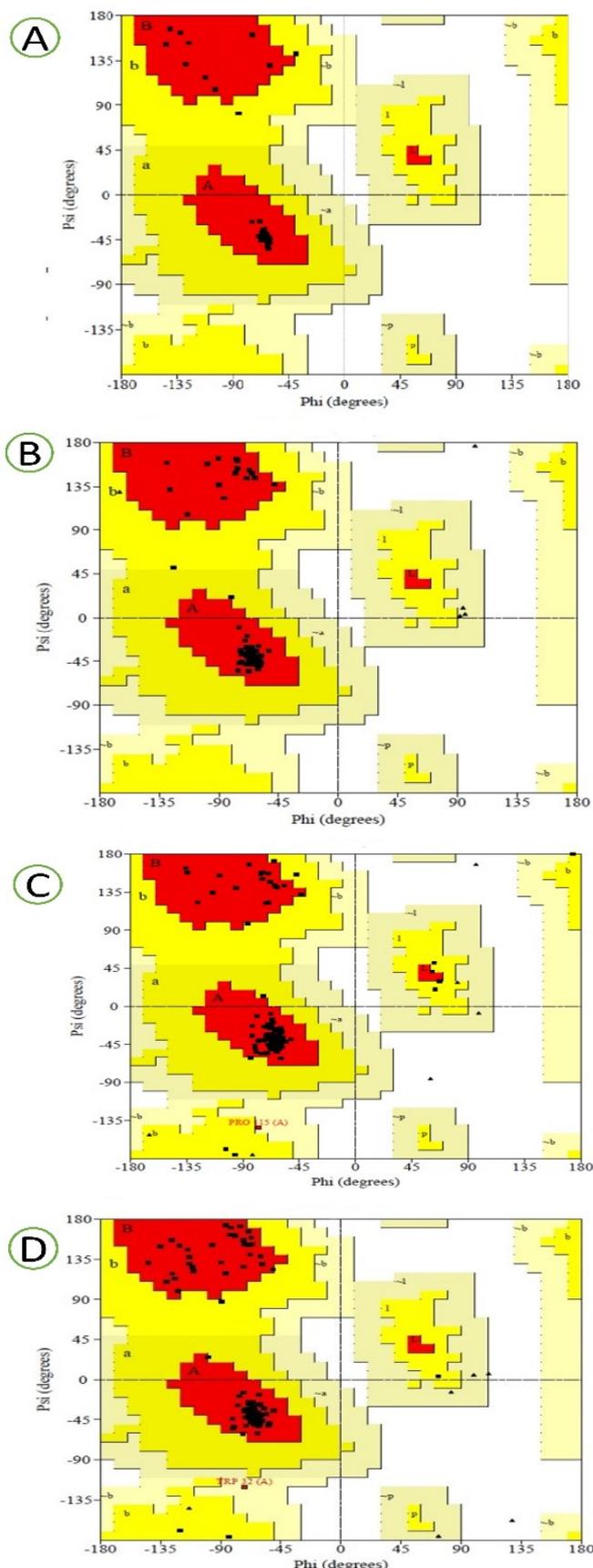


**Fig. 1.** Modelled 3D Structures of cytochrome proteins of *C. zedoaria*. (A) Cytochrome b559 subunit beta (B) Cytochrome b559 subunit alpha (C) Cytochrome b6 (D) Cytochrome b6-f complex subunit 4 (E) Cytochrome b6-f complex subunit 8 (F) Cytochrome b6-f complex subunit 5 (G) Cytochrome C Biogenesis protein.

ient way to view the distribution of torsion angles in a protein structure. QMEAN and ProSA servers were used for the quality estimation of predicted models. The Ramachandran plot for psbF showed that out of 39 residues, 33 were non-glycine and non-proline residues from which 31 (93.9%) residues were found to be in the most favoured region, 2 (6.1%) in additionally allowed region, 0 (0.00%) in generously allowed region and 0 (0.00%) in disallowed region shown in Fig. 2A and Table 4. In addition, 2 numbers of end residues, 2 numbers of glycine residues and 2 proline residues were found. For psbE, 69 (98.6%) residues
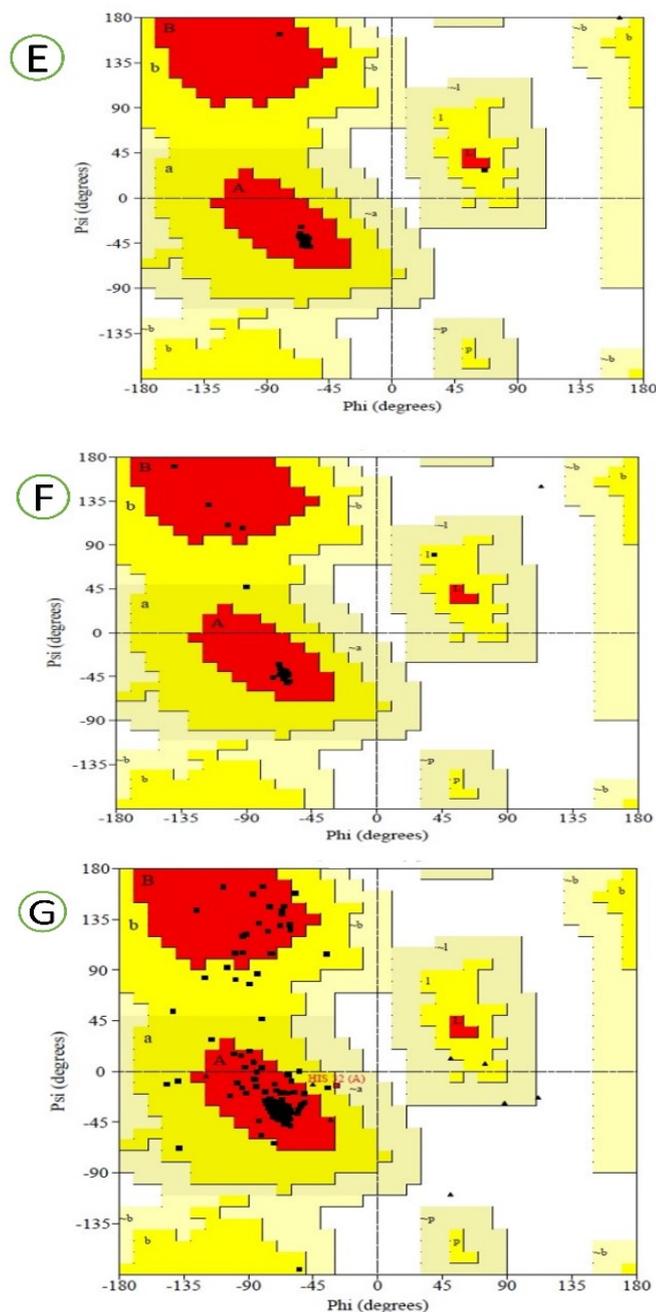
**Fig. 2.** Ramachandran plots of the predicted cytochrome proteins. (A) Cytochrome b559 subunit beta (B) Cytochrome b559 subunit alpha (C) Cytochrome b6 (D)Cytochrome b6-f complex subunit 4 (E) Cytochrome b6-f complex subunit 8 (F) Cytochrome b6-f complex subunit 5 (G) Cytochrome C Biogenesis protein.

were found in the most favoured region, 1 (1.4%) in additionally allowed region, 0 (0.0%) in generously allowed region, 0 (0.00%) in disallowed region shown in Fig. 2B and Table 4. For petB, 176 (94.1%) residues found in the most favoured region, 11 (5.9%) in additionally allowed region, 0 (0.0%) in generously allowed region, 0 (0.00%) in disallowed region shown in Fig. 2C and Table 4. For petD, 123 (95.3%) residues were found in the most favoured region, 5 (3.9%) in additionally allowed region, 1 (0.8%) in generously allowed region, 0 (0.00%) in disallowed region shown in Fig. 2D and Table 4. For petN, 24 (96.0%) residues found in the most favoured region, 1 (4.0%) in additionally allowed region, 0 (0.0%) in generously allowed region, 0 (0.0%) in disallowed region (Fig. 2E, Table 4). For petG, 28 (93.3%) residues were grouped in the most favoured region, 2 (6.7%) in additionally allowed region, 0 (0.0%) in generously allowed region, 0 (0.0%) in disallowed region shown in Fig. 2F and Table 4. For ccsA, 285 (94.1%) residues found in the most favoured region, 17 (5.6%) in additionally allowed region, 1 (0.3%) in generously allowed region, 0 (0.0%) in disallowed region shown in Fig. 2G and Table 4.

The QMEAN web server was used for determination of the model quality for psbF, psbE, petB, petD, petN, petG and ccsA. All the models were with good overall quality, as indicated by the QMEAN4 global score and Z-score. QMEAN Z-score with negative score shows low quality model whereas QMEAN4 score of above zero shows high quality model (20). The web server ProSA tool was used to evaluate each model's overall grade, Z-score reflected these proteins in comparison to all other structures estimated with X-ray crystallography and NMR. The psbF, psbE, petB, petD, petN, petG and ccsA have QMEAN4 values of -2.04, -1.20, -3.01, -1.57, -2.11, -1.74 and -12.87 respectively indicating good model quality. Fig. 3 showed QMEAN4 values of the proteins.

The Z-score of psbF, psbE, petB, petD, petN, petG and ccsA were 0.5, -0.83, -1.5, -0.58, -0.02, 0.14 and -3.73 respectively which indicated model's high quality. ProSA results of the proteins were shown in Fig. 4. The modelled proteins were also submitted to PMDB and accession numbers were assigned as PM0083506 (psbF), PM0083507

**Table 4.** Ramachandran plot statistics of cytochrome proteins in *C. zedoaria*

| Ramachandran plot parameters | Cytochrome b559 subunit beta | Cytochrome b559 subunit alpha | Cytochrome b6 | Cytochrome b6-f complex subunit 4 | Cytochrome b6-f complex subunit 8 | Cytochrome b6-f complex subunit 5 | Cytochrome C Biogenesis protein |
|---|---|---|---|---|---|---|---|
| Residues in the most favoured region | 31 | 69 | 176 | 123 | 24 | 28 | 285 |
| Residues additionally allowed region | 2 | 1 | 11 | 5 | 1 | 2 | 17 |
| Residues generously allowed region | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| Residues in disallowed region | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Total number of non-glycine and non-proline residues | 33 | 70 | 187 | 129 | 25 | 30 | 303 |
| Numbers of end residues (Excluding Glycine and Proline) | 2 | 2 | 2 | 2 | 2 | 2 | 2 |

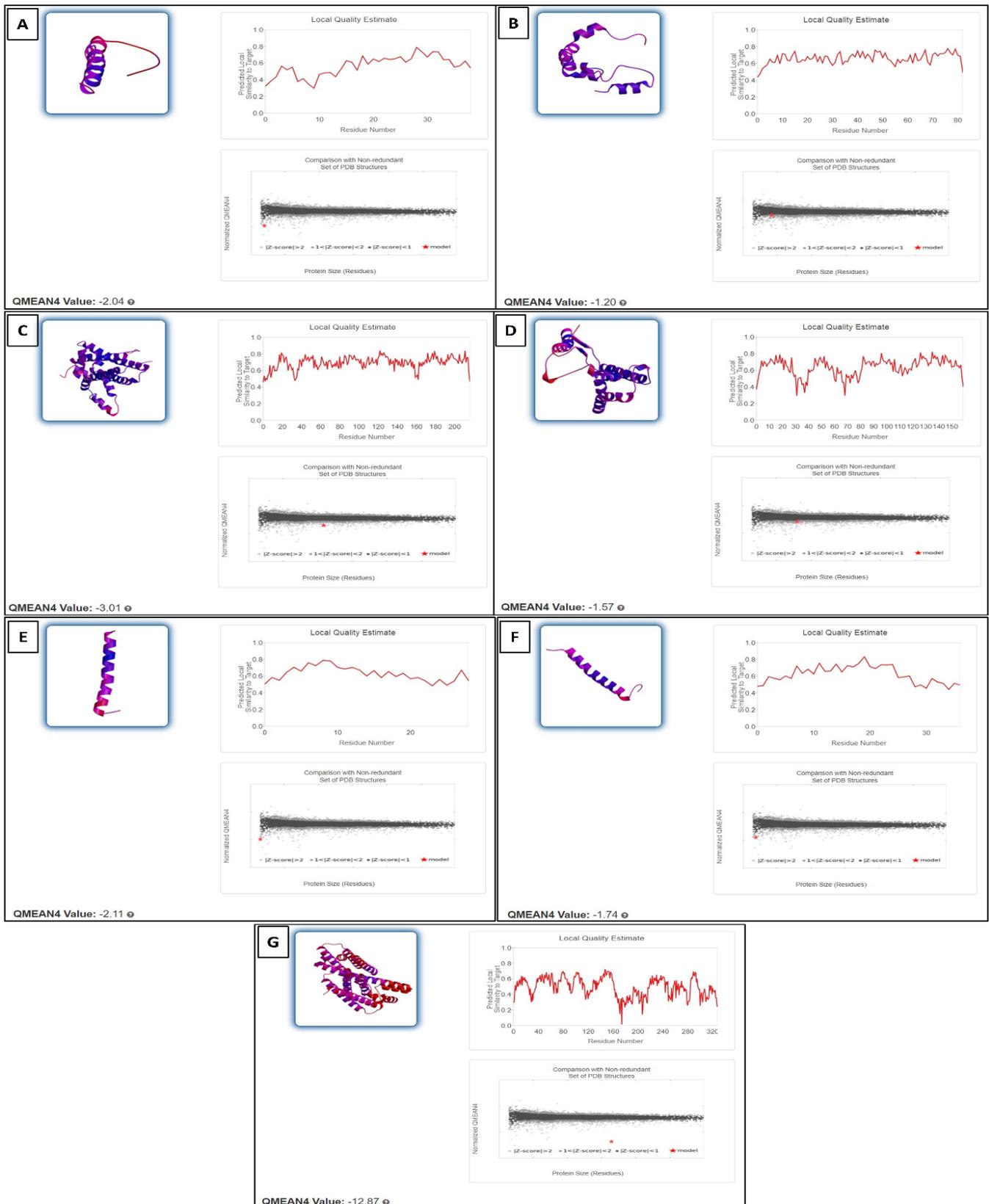| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Numbers of Glycine residues | 2 | 7 | 17 | 13 | 2 | 4 | 20 |
| Numbers of Proline residues | 2 | 4 | 11 | 16 | 0 | 1 | 4 |
| Total number of residues | 39 | 83 | 217 | 160 | 29 | 37 | 329 |



**Fig. 3.** QMEAN score plots for validation of cytochrome proteins. (A) Cytochrome b559 subunit beta (B) Cytochrome b559 subunit alpha (C) Cytochrome b6 (D) Cytochrome b6-f complex subunit 4 (E) Cytochrome b6-f complex subunit 8 (F) Cytochrome b6-f complex subunit 5 (G) Cytochrome C Biogenesis protein.

**Fig. 4.** Z-score plots of cytochrome proteins obtained using ProSA server. (A) Cytochrome b559 subunit beta (B) Cytochrome b559 subunit alpha (C) Cytochrome b6 (D) Cytochrome b6-f complex subunit 4 (E) Cytochrome b6-f complex subunit 8 (F) Cytochrome b6-f complex subunit 5 (G) Cytochrome C Biogenesis protein.
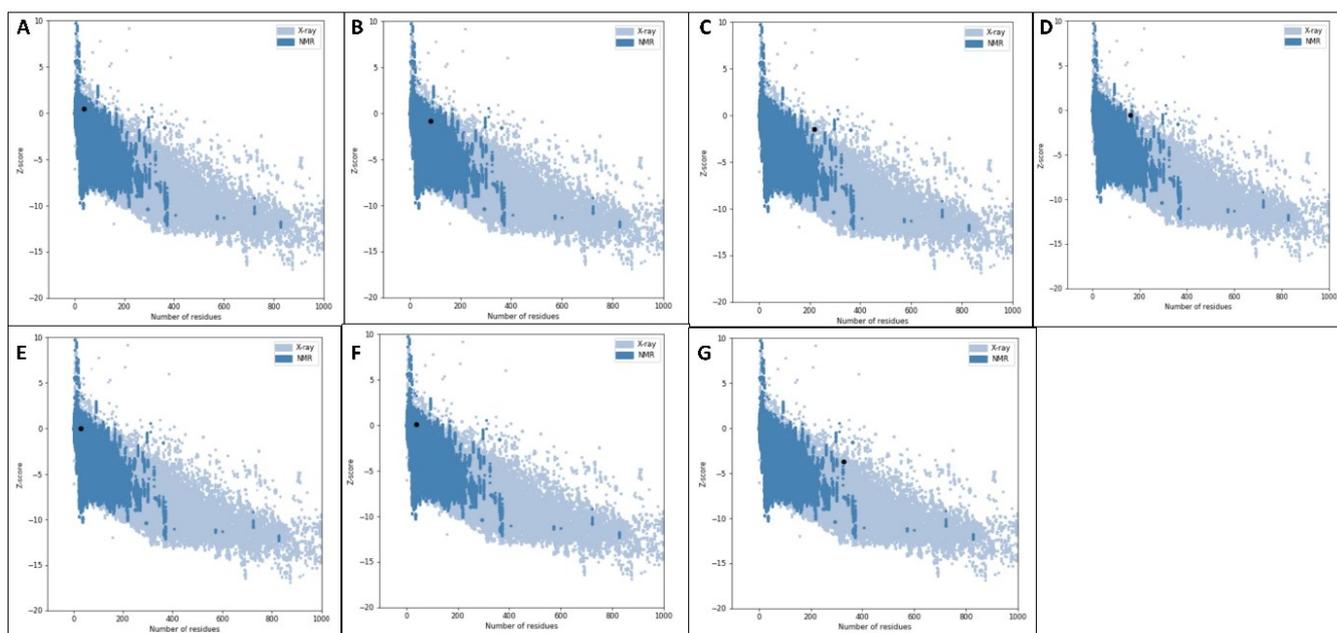
(psbE), PM0083508 (petB), PM0083509 (petD), PM0083510 (petN), PM0083511 (petG) and PM0083512 (ccsA).

## Discussion

*C. zedoaria* is having enough medicinal properties mainly due to presence of secondary metabolites and they were synthesized through several pathways. Cytochrome complex proteins play vital role to produce the secondary metabolites. Also, these proteins have important role in ATP formation, cyclic electron flow, biogenesis of c-type cytochrome of this plant. Most of the available protein sequences have not yet been structurally analysed in *C. zedoaria* through *in-silico* approach except few proteins like CzR1 gene (22, 23). Seven such complex proteins like psbF, psbE, petB, petD, petN, petG and ccsA have been targeted in the current study to predict their structures, which are not available in structural databases. As the function of a protein is mainly depends on its structure, it is very much important to predict the structure of protein before further validation. In this current study, the structures of the seven cytochrome proteins have been modelled through *in-silico* approach. The protein sequences were retrieved from Uni-Prot database. From ProtParam server of Expasy, it was found that 4 proteins namely psbE, petD, petN and petG are having pI less than 7, indicating the acidic nature along with another 3 proteins namely psbF, petB and ccsA are having pI greater than 7 indicating the basic nature (Table 2). At pI, proteins are stable and compact (24). Five proteins namely psbF, petB, petN, petG and ccsA showed II smaller than 40, which indicated that the proteins were stable. Range of GRAVY of *C. zedoaria* proteins were 0.076 - 1.545. The lowest value of GRAVY indicating the possibility of better interaction with water molecules (25) and here psbE was having

the lowest GRAVY value. Then the SOPMA server had been used for predicting 2D-Structures of the selected proteins. It was found that the percentage of alpha helices and extended beta strands were dominating over all other parameters. It defines that the proteins are highly stable in nature because these two parameters are highly responsible for the formation of hydrogen bonds leading towards the authentication of the rigidity of the protein structures (26). Then, best templates were selected from blastp results for 3D structures building. *In-silico* modelling had been done using MODELLER 9.23 version in advance modelling techniques. The best structures were selected based on DOPE score. Further, for structural validation and Ramachandran plot verification, PROCHECK server has been used. It was observed that, the residues of all the amino acids of the modelled proteins perfectly fit in the required allowed regions of the Ramachandran plot which indicated the better quality of the proteins (27). Then QMEAN server had been used for getting the QMEAN4 scores in which red star indicated the position of the modelled proteins related to the experimental structures of similar size. The scores nearer to zero value specify superior quality of the modelled protein structures. For each residue of the model, the 'local quality' was represented in x-axis and expected similarity to the native structure in y-axis. Usually, residues showing a score above 0.6 were considered to be high quality model (20). Finally, the ProSA server results gave the statistical values for native structures leading to confirmation and overall quality of the models. All the Z-score of the modelled proteins were found to be within the native conformation range. The black dot in the ProSA analysis in Fig. 4 was showing the position of the modelled protein with comparison to all other reported protein structures determined by X-ray crystallography and NMR (28). The ccsA protein along

with petB fallen in X-ray regions (light blue region of Fig. 4) and psbF, psbE, petD, petN along with petG were found in the NMR regions (deep blue region of Fig. 4). These results indicated that, the overall quality of the predicted models of proteins were quite good and could be helpful for further characterisation.

## Conclusion

The current study was focused on cytochrome proteins of *C. zedoaria* (Christm.) Roscoe. Various tools and servers have been implemented and the 3D structures of seven cytochrome complex proteins have been predicted and validated. These obtained 3D structures might be helpful in further investigations like determining the crystal structure of these hypothetical proteins, structural motifs, physiochemical properties, protein-protein interaction studies of various cytochrome proteins as well as function and mechanism of protein action. *C. zedoaria* is a highly endangered plant with immense medicinal properties and is needed to be analysed extensively for its future improvement.

## Acknowledgements

## Authors contributions

CST did data collection, writing and bioinformatics analysis, BCS helped in data collection and writing, MD helped in data collection and writing, DS helped in writing and manuscript preparation, SS helped in analysis, manuscript formatting and work design, BK designed the work, manuscript correction and bioinformatic analysis.

## Compliance with ethical standards

**Conflict of interest**: Authors do not have any conflict of interests to declare.

**Ethical issues**: None

## References

1. Lobo R, Prabhu KS, Shirwaikar A, Shirwaikar A. *Curcuma zedoaria* Rosc. (white turmeric): a review of its chemical, pharmacological and ethnomedicinal properties. J Pharm Pharmacol. 2009;61(1):13-21. https://doi.org/10.1211/jpp/61.01.0003

2. Etoh H, Kondoh T, Yoshioka N, Sugiyama K, Ishikawa H, Tanaka H. 9-Oxo-neoprocurcumenol from *Curcuma aromatica* (Zingiberaceae) as an attachment inhibitor against the blue mussel, *Mytilus edulis galloprovincialis*. Bioscience, Biotechnology and Biochemistry. 2003;67(4):911-13. https://doi.org/10.1271/bbb.67.911

3. Wilson B, Abraham G, Manju VS, Mathew M, Vimala B, Sundaresan S, Nambisan B. Antimicrobial activity of *Curcuma zedoaria* and *Curcuma malabarica* tubers. Journal of Ethnophar-macology. 2005;99(1):147-51. https://doi.org/10.1016/j.jep.2005.02.004

4. Dhal YO, Deo BA, Sahu RK. Antioxidant activity of enzymatic extracts of *Curcuma zedoaria* (Christm.). International Journal of Pharmacy and Pharmaceutical Sciences. 2012;4:343-46.

5. Chu HA, Chiu YF. The roles of cytochrome b559 in assembly and photoprotection of photosystem II revealed by site-directed mutagenesis studies. Frontiers In Plant Science. 2016;6:1261. https://doi.org/10.3389/fpls.2015.01261

6. Kurisu G, Zhang H, Smith JL, Cramer WA. Structure of the cytochrome b6f complex of oxygenic photosynthesis: tuning the cavity. Science. 2003;302(5647):1009-14. https://doi.org/10.1126/science.1090165

7. Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin MJ. UniProt: the universal protein knowledgebase. Nucleic Acids Research. 2004;32(suppl_1):D115-19. https://doi.org/10.1093/nar/gkh131

8. Santhoshkumar R, Yusuf A. In silico structural modeling and analysis of physicochemical properties of curcumin synthase (CURS1, CURS2 and CURS3) proteins of *Curcuma longa*. Journal of Genetic Engineering and Biotechnology. 2020;18(1):1-9. https://doi.org/10.1186/s43141-020-00041-x

9. Bharathi SV. *In silico* analysis of proteins of *Curcuma caesia* Roxb. Int J Phar Pharmaceut Sci. 2014;6(2):216-20.

10. Geourjon C, Deleage G. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Bioinformatics. 1995;11(6):681-84. https://doi.org/10.1093/bioinformatics/11.6.681

11. Webb B, Sali A. Comparative protein structure modeling using MODELLER. Current Protocols in Bioinformatics. 2016;54(1):5-6. https://doi.org/10.1002/cpbi.3

12. Fiser A, Do RK. Modeling of loops in protein structures. Protein Science. 2000;9(9):1753-73. https://doi.org/10.1110/ps.9.9.1753

13. Benkert P, Biasini M, Schwede T. Toward the estimation of the absolute quality of individual protein structure models. Bioinformatics. 2011;27(3):343-50. https://doi.org/10.1093/bioinformatics/btq662

14. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Research. 2007;35(suppl_2):W407-10. https://doi.org/10.1093/nar/gkm290

15. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Research. 2003;31(13):3784-88. https://doi.org/10.1093/nar/gkg563

16. Syngai GG, Barman PR, Bharali RU, Dey SU. BLAST: An introductory tool for students to Bioinformatics Applications. Keanean Journal of Science. 2013;2:67-76.

17. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The protein data bank. Nucleic Acids Research. 2000;28(1):235-42. https://doi.org/10.1093/nar/28.1.235

18. Wang S, Li W, Liu S, Xu J. RaptorX-Property: a web server for protein structure property prediction. Nucleic Acids Research. 2016;44(W1):W430-35. https://doi.org/10.1093/nar/gkw306

19. Rodriguez R, Chinea G, Lopez N, Pons T, Vriend G. Homology modeling, model and software evaluation: three related resources. Bioinformatics (Oxford, England). 1998;14(6):523-28. https://doi.org/10.1093/bioinformatics/14.6.523

20. Benkert P, Künzli M, Schwede T. QMEAN server for protein model quality estimation. Nucleic Acids Research. 2009;37(suppl_2):W510-14. https://doi.org/10.1093/nar/gkp322

21. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of pro-

teins. Nucleic Acids Research. 2007;35(suppl_2):W407-10. https://doi.org/10.1093/nar/gkm290

22. Kar B, Nanda S, Nayak PK, Nayak S, Joshi RK. Molecular characterization and functional analysis of CzR1, a coiled-coil-nucleotide-binding-site-leucine-rich repeat R-gene from *Curcuma zedoaria* Loeb. that confers resistance to *Pythium aphanidermatum*. Physiological and Molecular Plant Pathology. 2013;83:59-68. https://doi.org/10.6026/97320630009560

23. Joshi RK, Nanda S, Rout E, Kar B, Naik PK, Nayak S. Molecular modeling and docking characterization of CzR1, a CC-NBS-LRR R-gene from *Curcuma zedoaria* Loeb. that confers resistance to *Pythium aphanidermatum*. Bioinformation. 2013;9(11):560. https://doi.org/10.6026/97320630009560

24. Venkatasamy K. *In silico* Analysis and Homology Modeling of Putative Hypothetical Protein Q4QH83 of *Leishmania major*. Advanced BioTech. 2013;13(3):1-4.

25. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. Journal of Molecular Biology. 1982;157(1):105-32. https://doi.org/10.1016/0022-2836(82)90515-0

26. Perticaroli S, Nickels JD, Ehlers G, Sokolov AP. Rigidity, secondary structure and the universality of the boson peak in proteins. Biophysical Journal. 2014;106(12):2667-74. https://doi.org/10.1016/j.bpj.2014.05.009

27. Kumar P, Arya A. Ramachandran Plot: A simplified approach. 2018.

28. Srivani G, Behera SK, Dariya B, Chalikonda G, Alam A, Nagaraju GP. HIF-1α and RKIP: A computational approach for pancreatic cancer therapy. Molecular and Cellular Biochemistry. 2020;472(1):95-103. https://doi.org/10.1007/s11010-020-03788-6

§§§