ACCSCIENCE PUBLISHING

Gene & Protein in Disease

ORIGINAL RESEARCH ARTICLE

Structural, functional, phylogenetic, and molecular dynamic simulation study of PEST-containing nuclear protein: An e-science view

Nazeer Hussain Khan[1,2], Muhammad Shahid[3], Wenkang Wang[4], Saadullah Khatak[1,2], Ebenezeri Erasto Ngowi[1,5], Salma S. Mahmoud[1,6], Hao-Jie Chen[1], Lei Qian[1], Yangzhe Qin[1], Tao Li[1], Muhammad Zubair[7], Shazrul Fazry[3], Dong-Dong Wu[1,8]*, Chun Yang Zhang[9,10]*, and Xin-Ying Ji[1,11]*

[1]Henan International Joint Laboratory for Nuclear Protein Regulation, School of Basic Medical Sciences, Henan University, Kaifeng, Henan, China

[2]School of Life Sciences, Henan University, Kaifeng, Henan, China

[3]Department of Biological Sciences and Biotechnology, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, Bangi, Selangor, Malaysia

[4]Department of Breast Surgery, The First Affiliated Hospital, Zhengzhou University, Zhengzhou, Henan, China

[5]Department of Biological Sciences, Faculty of Science, Dar es Salaam University College of Education, Dar es Salaam, Tanzania

[6]Department of Medical Biochemistry and Molecular Biology, Faculty of Medicine, Ain Shams University, Cairo, Egypt

[7]Department of Wildlife and Ecology, University of Veterinary and Animal Sciences, Pattoki Campus, Pakistan

[8]School of Stomatology, Henan University, Kaifeng, Henan, China

[9]Department of Thoracic Surgery, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, Henan, P.R. China

[10]Department of General Thoracic Surgery, Hami Central Hospital, Hami, Xinjiang, P.R. China - Institute of Nursing and Health, Henan University, Kaifeng, Henan, China

[11]Kaifeng Key Laboratory of Infection and Biological Safety, Henan University College of Medicine, Kaifeng, Henan, China

***Corresponding authors:**
Dong-Dong Wu
(ddwubiomed2010@163.com)
Chun Yang Zhang
(zcy198200@163.com)
Xin-Ying Ji
(10190096@vip.henu.edu.cn)

Citation: Khan NH, Shahid M, Wang W, *et al*. 2022, Structural, functional, phylogenetic, and molecular dynamic simulation study of PEST-containing nuclear protein: An e-science view. *Gene Protein Dis*, 1(1):65.
https://doi.org/10.36922/an.v1i1.65

Received: April 12, 2022
Accepted: May 19, 2022
Published Online: June 3, 2022

Abstract

PEST-containing nuclear protein (PCNP) is a short-lived novel nuclear protein. It has been well evaluated that PCNP mediates the progression of several cancers, but the exact mechanisms are still under investigation. In this study, we provided an e-science view of PCNP protein from the aspects of protein structure, interactions, and bioinformatics-based analysis related to evolutionary features as well as proteomic profile. The phylogenetic relationship results reveal that PCNP is closely related to Pan troglodytes and the Bovidae family, while being distantly related to the Muridae family. The analysis of the physicochemical properties of PCNP demonstrated that it is a thermolabile protein which is slightly acidic and hydrophilic in nature. Further, coexpression and protein-protein interaction analyses were carried out, which demonstrated that the *PCNP* gene was remarkably expressed with *MORF4LI* and *RSL24D1* genes and has close interactions with TRAM1, PSMC6, SRP9, PRKRIR, UHRF2, and BMI1 proteins. Gene ontology and pathway enrichment analyses showed that PCNP has a high tendency to work in cell cycle regulation. Moreover, among the four 3D structure generating tools, I-TASSER-generated structure had the highest

quality factor score. The validation analysis revealed that the I-TASSER-generated structure exhibited the best quality factor score with maximum amino acids in the favored region. In addition, molecular dynamic simulation analysis approved the stable structure of the PCNP. This is the first study that highlights the usefulness of the understanding of the structural and functional analysis of the PCNP, which lays the groundwork for further experimental studies to validate the outcome.

Keywords: 3D structure; Molecular dynamic simulation; PEST-containing nuclear protein; Phylogeny; Physiochemical properties; Protein interactions

1. Introduction

Nuclear proteins (NPs) play a key role in the continuity of cell life by regulating the cell cycle and maintaining the complex organization of the cell[1]. There are various known families of NPs that play very essential roles in various cellular functions. With unique structure and functions, a family of NPs known as PEST-proteins (PEST-NPs), which are rich in proline (P), glutamic acid (E), serine (S), and threonine (T) sequences, acts as signal peptides for protein degradation[2,3]. PEST-NPs are also regarded as the guardians of the cell and are believed to play crucial roles in cell cycle regulation, ubiquitin proteasome pathway, glycosylation of nuclear pores, hexosamine biosynthetic pathways, cell nutrient uptake and cellular metabolism regulation, and nucleocytoplasmic transport[4-6]. It has been well reported that PEST sequences serve as a proteolytic signal to degrade the target proteins through proteasome or calpain proteolysis pathways[7,8]. In 2002, Mori *et al.* discovered a novel PEST-containing nuclear protein (PCNP), which is involved in vital cellular processes such as cell proliferation and mediates tumorigenesis. It is a short-living and small NP of only 178 amino acids with two remarkable PEST sequences[9]. Since its discovery, its role has been studied in several cancers for its relation with cancer progression, and now, it is recognized as an oncogenic protein.

PCNP has not been fully explored yet. There are very few molecular and genomic profiling studies regarding PCNP expression in cancer cells. However, in our previous *in vitro* and *in vivo* studies, we have exploited the tumorigenic roles of PCNP in neuroblastoma, lungs, and ovarian cancers[10-12]. It has been found that PCNP is upregulated in human lung adenocarcinoma, and mediates the proliferation, migration, and invasion of cells. However, knockdown of the *PCNP* gene is able to reverse the result and retard the progression of adenocarcinoma in the lungs[10]. In another study, we evaluated the PCNP gene expression in neuroblastoma using the SH-SY5Y and SK-N-SH cell lines. The findings showed that compared to the normal control group, the high expression of PCNP decreased the growth, migration, and invasion of human neuroblastoma through MAPK and PI3K/AKT/mTOR signaling pathways and promoted the apoptosis of neuroblastoma. Meanwhile, we also observed opposite results after silencing the *PCNP* gene in human neuroblastoma cells. These results were further strengthened by xenograft tumor analysis in animal experiments which employed both PCNP overexpression and knockdown approaches[11]. Recently, we also revealed a novel association of PCNP with the Wnt pathway[12]. PCNP is overexpressed in both ovarian cancer tissues and cells than in para-cancerous tissues and ovarian epithelial cells (IOSE80). Through *in vitro* and *in vivo* experiments on ovarian cancer cells (SK-OV-3 and A2780), it has been found that high expression of PCNP promotes growth, migration, and invasion of ovarian cancer cells and inhibits their apoptosis. Furthermore, PCNP could bind to β-catenin, promote β-catenin nuclear translocation, and then activate the Wnt/β-catenin signal pathway[12]. Based on all these findings, we justify that PCNP can be a promising biomarker for diagnosis and prognosis of cancer. Moreover, a potent PCNP inhibitor strategy could be adopted for exploring PCNP's role in tumorigenesis.

Therefore, it is essential to explore the mechanism of PCNP in cancer progression and its applicability as a potential target in cancer diagnosis and treatment. To the best of our knowledge, there has been no reported study on PCNP structure, interactions, and other bioinformatics analysis related to its evolutionary features as well as proteomic profile, and this is the first study to present the three-dimensional (3D) structural modeling, interaction with other proteins, and physiochemical properties of PCNP. The *in silico* analysis performed in this study demonstrated the features and properties of PCNP collected from different online web servers and software, which help increase our understanding of PCNP and its functions in cellular activities in tumorigenesis.

2. Materials and methods

2.1. Data retrieval and multiple sequence alignment

The protein ID and FASTA amino acid sequence of PCNP were retrieved from the National Center for Biotechnology

Information (NCBI) (https://www.ncbi.nlm.nih.gov/) with accession reference NP_065090.1. To accomplish this study, we employed various bioinformatics tools and software as listed in Table 1.

2.2. 3D structure prediction

PCNP has no reported structure in the protein data bank (PDB) database (https://www.rcsb.org)[13]. Therefore, to generate the 3D structure, both protein comparative homology and sequence-based structure modeling approaches were utilized. We adopted four DeepMind learning tools that predict the protein structures, for example, AlphaFold (https://alphafold.ebi.ac.uk/), RoseTTAFold (https://boinc.bakerlab.org/rosetta/), Iterative Threading Assembly Refinement (I-TASSER) (https://zhanggroup.org/I-TASSER/), and Robetta (https://robetta.bakerlab.org/)[14-17]. AlphaFold and RoseTTAFold are new computational tools that use deep learning algorithms to generate protein structures. Besides, they predict the protein structures in cases where no homology structure is known by incorporating the limited physical and biological knowledge into the deep learning algorithms[18,19]. I-TASSER and Robetta are publicly available web-based tools that predict 3D protein structure. They use the amino acid sequence as input and generate the 3D structure either by searching for multiple threading alignments for the template from the PDB or using iterative structural assembly simulations, a *de novo* structure prediction method[16,17]. Graphical inspection and representation of the models were done using UCSF Chimera software version 1.14 (http://www.cgl.ucsf.edu/chimera/)[20].

Furthermore, the resultant structures from the aforementioned tools were further subjected to the ERRAT and PROCHECK tools. Both are in-built tools of the Structure Analysis and Verification Server (SAVES) meta-server version 6.0 (https://saves.mbi.ucla.edu/) that determine the quality factors (degree of appropriateness) and stereochemical evaluation of the query protein structure[21,22]. In addition, the quality model energy analysis (QMEAN) was performed using the QMEAN server accessed from https://swissmodel.expasy.org/qmean/. It is a composite scoring function that estimates the quality of a protein structure model by deriving the local (per residue) or global (for the entire protein structure) scores in terms of a Z-score[23].

2.3. Molecular dynamic simulation (MDS)

MDS of the best scored PCNP protein structure from the above-mentioned tools was carried out using GROMACS software[24]. In *in silico* studies, MDS is considered the key analysis to identify the interactions of protein molecules and their structural characteristics[25]. For MDS run, the PCNP

Table 1. List of various bioinformatics tools employed in the present study.

S. No.	Database/tool	Functions
1.	National Center for Biotechnology Information (https://www.ncbi.nlm.nih.gov/)	To retrieve amino acid sequence of protein
2.	Protein Data Bank (https://www.rcsb.org)	Structure of protein
3.	AlphaFold (https://alphafold.ebi.ac.uk/)	Protein structure construction
4.	RoseTTAFold (https://boinc.bakerlab.org/rosetta/)	Protein structure construction
5.	I-TASSER (Iterative Threading Assembly Refinement; https://zhanggroup.org/I-TASSER/)	Protein structure construction
6.	Robetta (https://robetta.bakerlab.org/)	Protein structure construction
7.	UCSF Chimera software version 1.14	Graphical inspection and representation of the models
8.	SAVES meta-server version 6.0 (https://saves.mbi.ucla.edu/)	To determine the quality factors of protein structure
9.	Quality Model Energy Analysis (QMEAN; https://swissmodel.expasy.org/qmean/)	Quality of a protein structure by deriving the local or global scores in terms of Z-score
10.	GROMACS software	Molecular dynamic simulation of protein structure
11.	ProtParam tool of ExPASy (http://web.expasy.org/protparam/)	To determine the basic physicochemical properties
12.	SOPMA (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html)	For the secondary structure prediction of a protein
13.	GeneMANIA server (https://genemania.org/)	To check the co-expression analysis of a protein
14.	STRING protein database (https://string-db.org/)	PPI of protein
15.	Cytoscape software version 3.8.2	To construct the coexpression and PPI networks
16.	Network-Analyst server (https://www.networkanalyst.ca/)	To assimilate the biological processes, molecular functions and cellular components, and to enrich pathway analysis of protein
17.	Enrichr web-server (https://maayanlab.cloud/Enrichr/)	To validate the Network Analyst results
18.	Molecular Evolutionary Genetic Analysis (MEGA Home; megasoftware.net)	To construct phylogenetic tree of gene
19.	Sequence Demarcation Tool Version 1.2 (SDTv1.2; http://web.cbio.uct.ac.za/~brejnev/)	For pairwise sequence identity of a gene in evolutionary analysis
20.	Google Scholar	To retrieve the literature

PPI: Protein-protein interaction

protein structure was prepared using the GROMOS96 43a1 force field solvated by the single point charge water model and was fixed in a periodic cubic solvated box. Other parameters were neutralized by adding the salt (NaCl) concentration (0.15 M), energy minimization for 5000 steps of the prepared system, and equilibration type NVT/NPT at 300 K and 1 bar of pressure. The final MDS step was processed for 50 ns of time. The MDS was analyzed by calculating the root mean square deviation (RMSD) and root mean square fluctuation (RMSF) values of the protein structure.

2.4. Physicochemical characterization and secondary structure prediction

The basic physicochemical properties, such as theoretical isoelectric point (pI), molecular weight, total number of positive and negative residues, extinction coefficient instability index, aliphatic index, and grand average hydropathy, were computed using the ProtParam tool of ExPASy (http://web.expasy.org/protparam/)[26-28]. For the secondary structure prediction of PCNP, SOPMA bioinformatics tool (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html) was used with its default query parameters[29]. This tool uses the self-optimized prediction method and a neural network method (PHD) to predict the secondary structure of proteins[29,30].

2.5. Coexpression and protein-protein interaction (PPI) analysis

The coexpression analysis of the PCNP was performed by the GeneMANIA server (https://genemania.org/) by setting *Homo sapiens* as the species organism and the maximum number of resultant genes and attributes as 10. It is a web interface that has a high-accuracy prediction algorithm and a large database for gene functional analysis, including coexpression, genetic interaction, colocalization, physical interaction, pathway interaction, and shared protein domains of the submitted query[31]. Furthermore, the PPI of PCNP was evaluated by accessing the search tool for the retrieval of interacting genes (STRING) protein database (https://string-db.org/). The PPI network was constructed based on the data from tandem affinity purification assay, affinity chromatography technology assay, and coimmunoprecipitation assay. The STRING database query terms were cutoff at 0.4 and the maximum additional interaction was 10[32]. Cytoscape software version 3.8.2 (https://cytoscape.org/) was employed to construct the coexpression and PPI networks[33].

2.6. Gene ontology (GO) and pathway enrichment analysis

To assimilate the biological data of GO, including biological processes, molecular functions, and cellular components, and enriched pathway analysis of PCNP, the Network Analyst server (https://www.networkanalyst.ca/) was employed[34,35]. It is a publicly available online tool widely used for gene expression profiling, transcriptional factor analysis, PPI networks, pathway analysis, toxicogenomics, and pharmacogenomics studies[36,37]. In addition, the Enrichr web server (https://maayanlab.cloud/Enrichr/) was utilized to validate the Network Analyst results[35].

2.7. Phylogenetic analysis

To determine the evolutionary history of PCNP between humans and other vertebrates, the phylogenetic tool, molecular evolutionary genetics analysis (MEGA) X, was used[38]. PCNP has no reported paralogs in the biological database (ENSEMBL) or in the available literature. Orthologs of PCNP were used for tree construction and sequences of eight orthologous candidate proteins were retrieved from the NCBI database. Alignment of multiple sequences was carried out using the MUSCLE program implemented within MEGA X. For pairwise sequence identity, Sequence Demarcation Tool Version 1.2 (SDTv1.2; http://web.cbio.uct.ac.za/~brejnev/) was used[39]. The best-fit substitution model was estimated in MEGA and selection was done based on the Neighbor-Joining Method. The rate of variation among sites was modeled with a gamma distribution (shape parameter = 1). All positions that contained alignment gaps and missing data were eliminated from the analysis. A phylogenetic tree was constructed using the unweighted pair group method with arithmetic mean, and the validity of the tree generated was tested by bootstrap analysis of 1000 pseudoreplicates[40,41].

3. Results and discussion

3.1. Structure prediction

Four tools were utilized to generate the 3D structure of the PCNP protein. From the results of all these tools, we selected the best score-generating model and subjected it to further validation through the ERRAT and PROCHECK servers. In addition, the quality model QMEAN score was evaluated for the best scoring models. The graphical illustration of these four best models is presented in Figure 1. Following the 3D model analysis, it was discovered that the I-TASSER generated model had the highest quality factor score among the four best-scoring models, as shown in Figure 1D and Table 2. Moreover, the I-TASSER model structure was subjected to MDS for profound validation analysis.

3.2. Molecular dynamics simulation

An MDS was conducted for the analysis of the stability and flexibility of the I-TASSER-generated PCNP protein model. The RMSD and RMSF of the protein structure

were calculated using MDS for the 50-ns time interval. The results of RMSD and RMSF are presented in Figure 2. The RMSD values were observed in the range of 0.2–0.9 nm of the protein model. In brief, the results revealed that the RMSD value came to equilibrium, maintaining around 0.6–0.8 nm throughout the MDS run of 50 ns (Figure 2A). Although the lack of experimental structure, these MDS results suggest that the protein model is dynamically stable throughout the MDS period with a slightly higher RMSD value. Similarly, the amino acid residue wise fluctuation, the RMSF value, was plotted and is shown in Figure 2B. The RMSF analysis demonstrated that notable deviations have occurred in the first 20–25 amino acid residues, but this type of high fluctuation was not observed in other residues throughout the MDS period.

3.3. Physicochemical characterization and secondary structure prediction

The ProtParam tool, available on the ExPASy server, predicts the physiochemical properties of the protein under investigation. The PCNP gene encodes a protein of 178 amino acids ($C_{805}H_{1304}N_{238}O_{280}S_4$) having a molecular weight of 18924.88 kDa. The theoretical isoelectric point (pI) of this protein was found to be 6.86 with 28 negatively charged residues (Asp + Glu) and 28 positively charged residues (Arg + Lys). At this computed pI, the PCNP protein seemed to be stable and compact and was considered slightly acidic. These results will be helpful in preparing a buffer system for purification by the pI method[42]. The instability index of PCNP protein was computed as 15.56, which indicates that the protein may be marginally stable as the values are lower than 40. The instability index of a protein greater than 40 indicates that the protein is likely to be unstable.

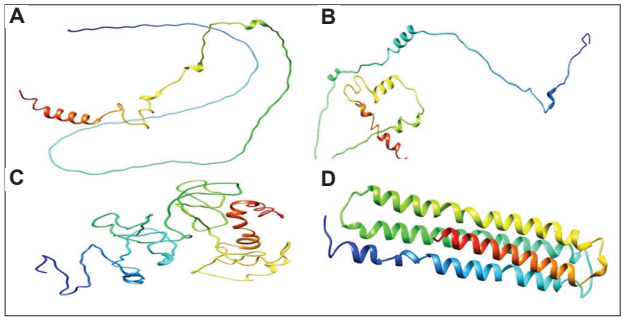

Figure 1. The 3D structures of PCNP protein predicted by (A) AlphaFold, (B) RoseTTAfold, (C) Robetta servers, and (D) I-TASSER.

In addition, the value of the aliphatic index, which is the relative volume of aliphatic side chains, indicates the thermal stability of a protein structure. The higher the positive factor value, the greater the thermal stability of the respective protein. The aliphatic index of PCNP was calculated to be 89.93, indicating its stability at higher temperatures[43]. The grand average of the hydropathicity of the protein was predicted to be −0.463, which indicates the hydrophilic nature of the protein and improved interaction with water. The detailed results are documented in Figure S1.

Of the total 178 amino acids, 58 (32.58%) were found in the alpha helix, 11 (6.18%) in the extended strand, 8 (4.49%) in the beta turns, and 101 (56.74%) in the random coil region of the PCNP protein as predicted by the SOPMA tool. These results demonstrate that the PCNP protein secondary structure has a dominant random coil region followed by beta turns and an alpha helix. The detailed results are presented in Figure S2.

3.4. GO and pathway enrichment analysis

GO is an efficient machine learning method useful for the identification of biological processes, molecular functions, and cellular components of a query protein. In the present study, a total of 46 biological processes, 10 molecular functions, and 10 cellular components of PCNP were identified by the Network Analyst server (Table S1). All the predicted biological processes results were below 0.05 (*P*-value) and considered a significant outcome. The top 10 biological processes results are demonstrated in Figure 3A. In biological processes analysis, our results indicate that PCNP has proximate and profound involvement in essential cellular activities, such as cell cycle, cycle arrest, and cellular catabolic processes (Table S1). Out of the 10

Table 2. The quality factors-score values of four structures of PCNP by predicted by various corresponding servers.

Protein	SERVER	SWISS MODEL		PROCHECK			PROSAWEB	ERRAT
PCNP		QMEAN4 Score	QMEANDisCo Global Score	Favored region (%)	Allowed region (%)	Disallowed region (%)	Z-score	ERRAT
1	AlphaFold2	−9.33	0.34±0.06	61.5	37.8	0.7	−1.71	92.4528
2	RoseTTAFold	−2.39	0.42±0.06	80.4	17.6	2	NA	91.3386
3	I-TASSER	−6.37	0.28±0.06	87.2	11.5	1.4	NA	94.7059
4	Robetta	−12.32	0.31±0.05	40.4	55.7	3.8	−3.08	82.8244

Based on these results, I-TASSER model was selected as best structure and subjected to further MDS analysis.

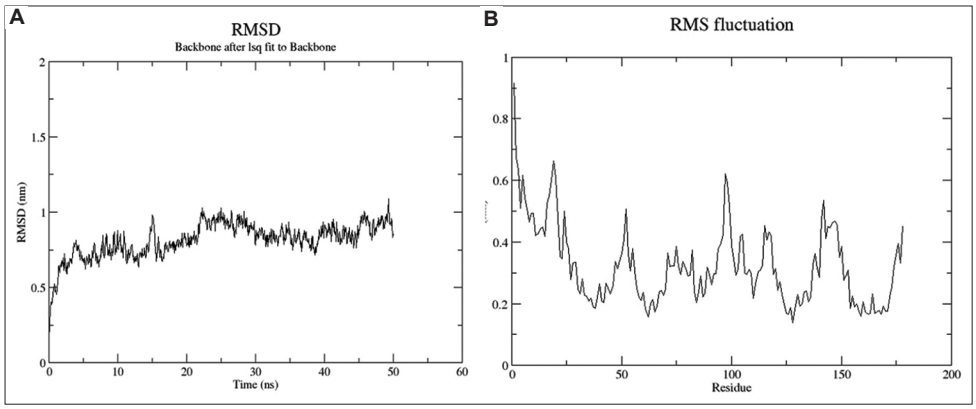

Figure 2. Molecular dynamic simulation of coordinate (A) root mean square deviation and (B) root mean square fluctuations of PCNP structure generated by I-TASSER.

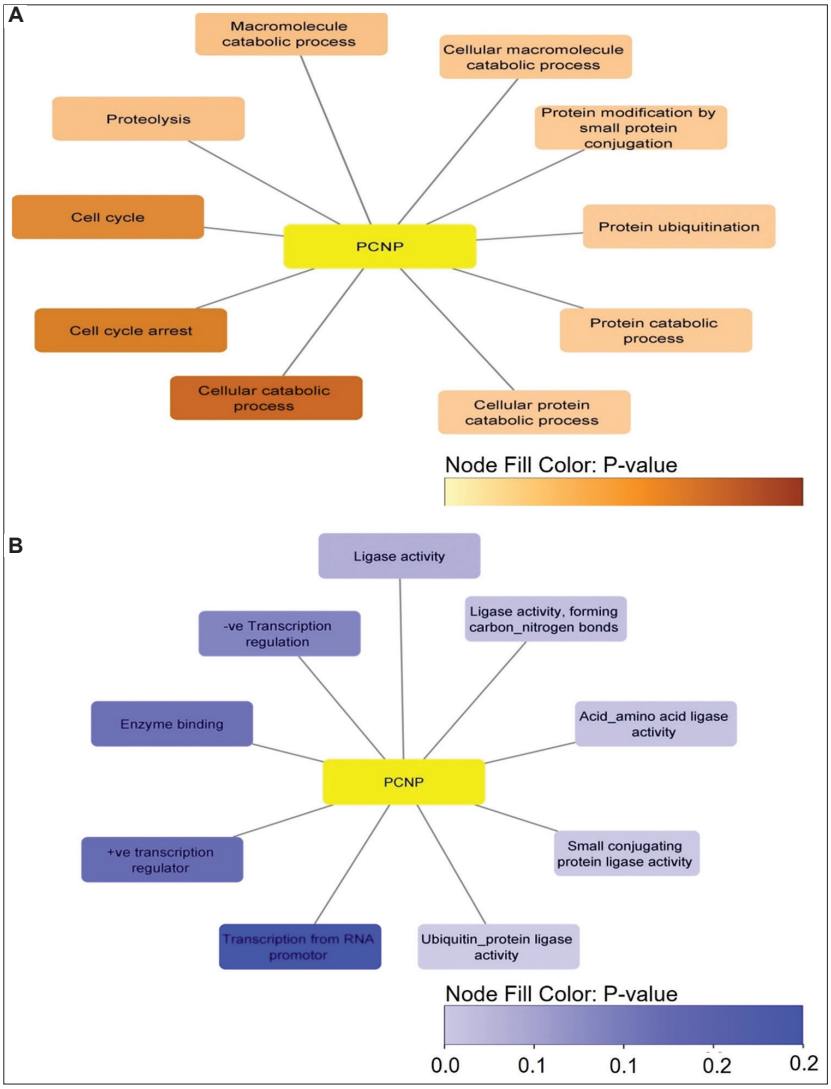

Figure 3. Gene ontology network of (A) the top 10 enriched biological processes and (B) the 10 enriched molecular functions of PCNP constructed by Cytoscape software. In the constructed networks, nodes are color-shaded according to their P-value score i.e., red > orange > yellow for biological processes, and dark blue > blue > light blue for molecular functions.

molecular functions, ubiquitin protein ligase activity, small conjugating protein ligase activity, amino acid ligase activity, and ligase activity forming carbon nitrogen bonds showed $P < 0.05$. The network of molecular functions constructed by Cytoscape software is presented in Figure 3B. Similarly, the cellular components results show the distribution of PCNP solely in the nucleus of the cell (Table S1). Furthermore, the pathway enrichment analysis was performed using the Network Analyst server. There were no results for the PCNP protein when searching broad queries on both servers (Network Analyst and Enrichr) for the Kyoto Encyclopedia of Genes and Genomes (KEGG) Reactome, and WikiPathways.

3.5. Coexpression and PPI analysis

Co-expression of PCNP was investigated by GeneMANIA server. The outcomes of GeneMANIA were subjected to Cytoscape software to construct the network based on their coexpression score. The coexpression results reveal that PCNP has significant expression with *MORF4LI* and *RSL24D1* genes, and the least association of coexpression with *ATF1*, *NFYB*, and *CMPK1* genes (Figure 4). *MORF4* (mortality factor on chromosome 4) is a senescence-inducing gene in tumor cells. The previous research has demonstrated that MORF4 protein is found both in the cytoplasm and the nucleus of cells[44]. In addition, RSL24D1 is a NP involved in ribosomal biogenesis[45]. Hence, these findings are the first to confirm that PCNP is a NP, and it might be a potent protein that plays a role in the regulation of tumor cell progression and ribosomal biogenesis through influencing the expression of *MORF4* and *RSL24D1* genes.

Moreover, in the present study, the PPI was determined based on the data of tandem affinity purification assay, affinity chromatography technology assay, and coimmunoprecipitation assay. From the results, four proteins, namely, TRAM1, PSMC6, SRP9, and PRKRIR, were extracted using tandem affinity purification and coimmunoprecipitation assays. Using the affinity chromatography technology assay, two proteins, UHRF2 and BMI1, were found to have an interaction with PCNP protein (Figure 5). TRAM1 (Translocating chain-associated membrane protein 1) and SRP9 (Signal Recognition Particle 9) are involved in the translocation of nascent secretory proteins by facilitating proper chain folding into or through the endoplasmic reticulum membrane[46]. To this end, it can be speculated that PCNP might be associated with endoplasmic reticulum protein transport by making direct or indirect interaction with TRAM1 or SRP9. In addition, PSMC6 is located in the base region of the proteasome 19S regulatory particle and is involved in the translocation of the substrates to the core particle. A recent study reported that

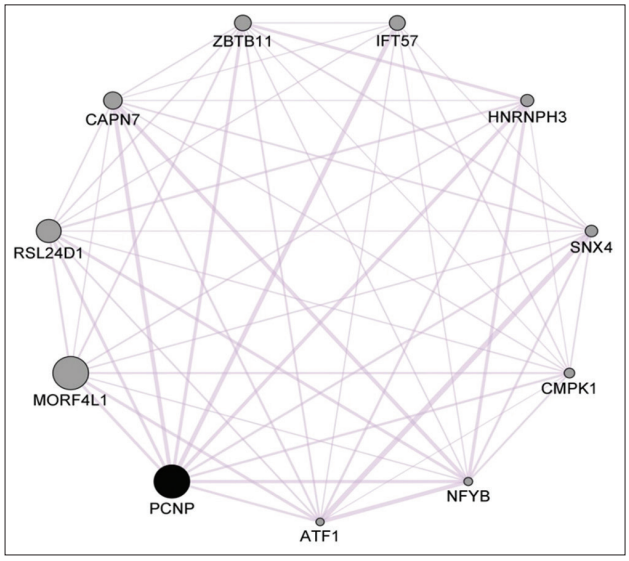

Figure 4. Coexpression network of 10 common targets of PCNP constructed with Cytoscape software. In the network, the size of the node corresponds to the coexpression affinity level with the PCNP. The bigger the size of the node, the greater the coexpression level with the respective gene.

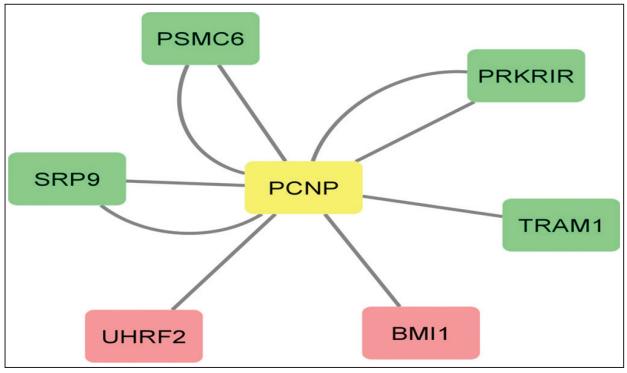

Figure 5. The protein-protein interaction network of PCNP protein based on the experimental models constructed with Cytoscape software. The proteins in the green color nodes were collected from the tandem affinity purification assay and coimmunoprecipitation assays, and the proteins in the pink color nodes were filtered from the affinity chromatography technology assay.

PSMC6 promotes the apoptosis of osteoblasts by halting the PI3K/AKT signal transduction pathway. Interestingly, a previous research study demonstrated PCNP activity through the PI3K/AKT signal transduction pathway. Therefore, these results suggest that the cross-talk between PCNP and PSMC6 is needed to regulate vital functions of the cell through the PI3K/AKT signal transduction pathway[47]. Moreover, UHRF2 is also a NP involved in cell cycle regulation. It halts cell proliferation and acts as a tumor suppressor factor. BMI1 is a member of polycomb-repressive complex 1 that is necessary for chromatin remodeling or silencing. Importantly, it also plays a role in the self-renewal of somatic stem cells[48]. Apart from these PPI findings,

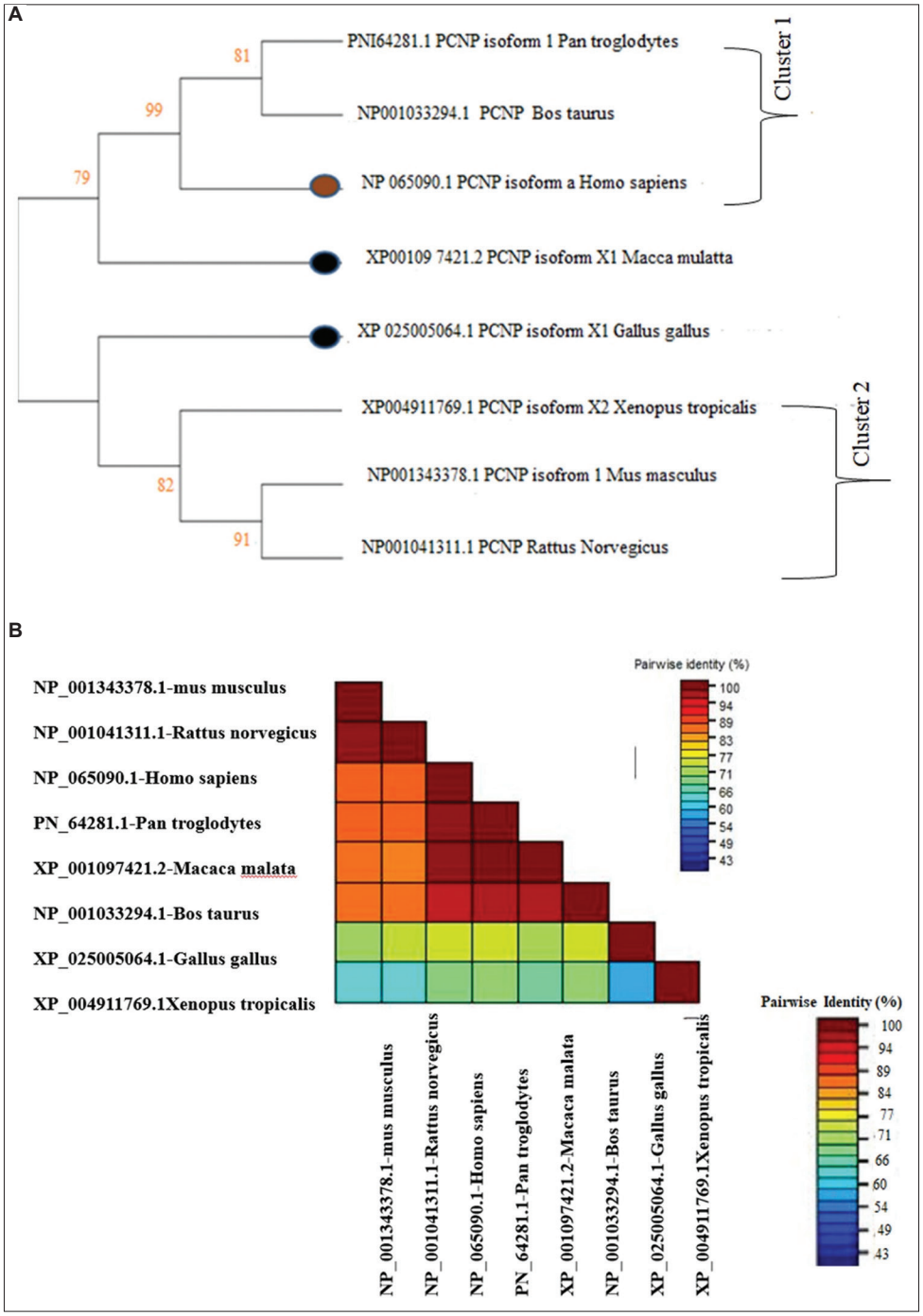

Figure 6. (A) Phylogenetic tree of PCNP) of human and other mammalian species. The tree was constructed using MEGA7 software by the Neighbor-Joining method. Different mammalian species along with GenPept IDs are shown. Numbers along the branches refer to the bootstrap value (percentage of confidence). PCNP sequence (NP 065090) used in this study is highlighted with red solid circle and out-grouped sequence is represented with black solid circle. (B) Pair-wise identity matrix of PCNP sequence of various mammalian species. A color-coded pair-wise identity matrix was generated from PCNP protein sequences (species name and GenPept IDs are shown) using SDT program. Each colored cell represents a percentage of identity score between two sequences (one indicated horizontally while the other vertically). Species forming different subgroups are demarcated by red box.

published literature on coexpression of PCNP reveals that dolichol-phosphate mannosyltransferase subunit 1 (DPM1), a mannose donor in various glycosylation reactions and a catalytic subunit of the polyol-mannose phosphate (DPM) synthesis complex, is coexpressed with PCNP at the post-translational modification (PTM) level in patients with

congenital disorders[49]. PTMs refer to amino acid side chain modifications in some proteins after their biosynthesis. Disruption in PTMs can cause malfunctioning of vital biological processes and, hence, lead to various diseases[50]. Therefore, considering the DPM1 and PCNP coexpression as synchronization of PCNP functions, there is an urgent need to predict PTMs in this least explored protein, which will provide a clear picture of PCNP functions, especially in cancer onset. Based on these already published findings and the predicted outcomes of the present study, it can be deduced that PCNP might have diverse functions, although many of which remain unreported.

3.6. Phylogenetic analysis

The phylogenetic neighbor-joining tree presented in this study reveals many interesting characteristics of the candidate protein in vertebrates. The human PCNP protein phylogenetic tree revealed remarkable similarities to *Pan troglodytes* and *Bos taurus*. The results indicate a close relationship with both species but that they have ancestor linkage in a distinguished clade. The generated phylogenetic tree of PCNP is presented in Figure 6A. Besides, the evolutionary tree shows that the ancestry of PCNP protein is closer to the Bovidae family but far away from other primates like *Mus musculus*. The tree also shows that in the PCNP phylogeny, *Gallus gallus* is an out-group. Moreover, based on the pairwise sequence identity, the PCNP protein from the Bovidae family could be clustered as one subgroup. The pair-wise identity matrix of PCNP sequences and its eight orthologs in various vertebrates is demarcated in Figure 6B. These findings additionally support the phylogenetic tree that presents an indication of the molecular kinship of PCNP proteins belonging to diverse mammalian species. Overall, the *in silico* results reveal many hidden functional, structural, and evolutionary characteristics of the human PCNP protein.

3.7. Scope and limitations of the study

Doubtlessly, the *in silico* study presented here provides a critical, quick, and low-cost structure prediction and functional elucidation of PCNP, and it will add more literature and understanding of PCNP for the future studies. However, there are some limitations to this study, such as the fact that it is a computational study. A study based on the results that were retrieved from public data-tools has no affirmation from a wet laboratory study. Results generated from software implementation of the data have high false positive or false negative rates, which restrict us from being fully confident in the results. Furthermore, PCNP is the least explored protein, so there is a lack of data to validate the molecular dynamics experiments and structural stability. Therefore, it is highly recommended to perform functional assays in cell lines as well as crystallographic analysis to fully understand the PCNP structure and its binding abilities.

4. Conclusion

Our study presents an *in silico* structural and functional interpretation of the PCNP by employing various bioinformatics tools. We identified the phylogenetic relationship of PCNP and then performed structural and functional analysis. Overall, PCNP is slightly acidic, hydrophilic in nature, and a thermostable protein. The protein secondary structure has a dominant random coil region followed by beta turns and an alpha helix. Through the coexpression and PPI analyses, we discovered that the *PCNP* gene is expressed with *MORF4LI* and *RSL24D1* genes and has close interactions with TRAM1, PSMC6, SRP9, PRKRIR, UHRF2, and BMI1 proteins. The GO results are consistent with the previously reported studies and endorse PCNP's involvement in vital cell functions, such as cell cycle and cancer progression as an oncogenic protein. Among the four different 3D structure generation tools, I-TASSER generated a model that had the highest quality factor score, which was further validated through MDS. To date, this is the first comprehensive study to elucidate the structural outlook and functional details of PCNP. This computational work, therefore, provides an essential perspective of PCNP, which can be useful for proteomics or genomic research involving extensive *in vivo* experimental work on this vital NP, and helps deepen our understanding regarding the function of PCNP in the nucleus and its potential role as a diagnostic and therapeutic target in cancer biology discipline.

Acknowledgments

We acknowledged all the database websites tools and servers for providing a platform of knowledge and data interpretation as well. We also thankfully acknowledged Mr. Abbas Khan for helping us with the editing of this manuscript. The authors thank the expert reviewers for taking their time to kindly review the manuscript and provide their valuable suggestions to improve the manuscript. Nazeer Hussain Khan pays countless thanks to his Lala for his presence and support in the journey of PhD.

Funding

This work was financially supported by the National Natural Science Foundation of China (No. 81670088 and No. 81602708).

Conflict of interest

All the authors declare no conflict of interest.

Author contributions

Conceptualization: Nazeer Hussain Khan and Wenkang Wang

Data curation: Nazeer Hussain Khan and Wenkang Wang

Formal analysis: Nazeer Hussain Khan, Wenkang Wang, Muhammad Shahid, Saadullah Khatak, and Ebenezeri Erasto Ngowi

Investigation: Muhammad Shahid, Saadullah Khatak, and Ebenezeri Erasto Ngowi

Methodology: Muhammad Shahid, Saadullah Khatak, and Ebenezeri Erasto Ngowi

Resources: Salma S. Mahmoud, Hao-Jie Chen, Lei Qian, Yangzhe Qin, Dong-Dong Wu, Chun Yang Zhang, and Xin-Ying Ji

Software: Salma S. Mahmoud, Hao-Jie Chen, Lei Qian, and Yangzhe Qin

Supervision: Dong-Dong Wu, Chun Yang Zhang, and Xin-Ying Ji

Visualization: Salma S. Mahmoud, Hao-Jie Chen, Lei Qian, and Yangzhe Qin

Writing – original draft: Nazeer Hussain Khan and Wenkang Wang

Writing – review and editing: Tao Li, Muhammad Zubair, and Shazrul Fazry

References

1. Sigal A, Milo R, Cohen A, *et al.*, 2006, Dynamic proteomics in individual human cells uncovers widespread cell-cycle dependence of nuclear proteins. *Nat Methods*, 3(7): 525–531.

   https://doi.org/10.1038/nmeth892

2. Afzal A, Sarfraz M, Li GL, *et al.*, 2019, Taking a holistic view of PEST-containing nuclear protein (PCNP) in cancer biology. *Cancer Med*, 8(14): 6335–6343.

   https://doi.org/10.1002/cam4.2465

3. Chevaillier P, 1993, Pest sequences in nuclear proteins. *Int J Biochem*, 25(4): 479–482.

4. Sarfraz M, Afzal A, Khattak S, *et al.*, 2021, Multifaceted behavior of PEST sequence enriched nuclear proteins in cancer biology and role in gene therapy. *J Cell Physiol*, 236(3): 1658–1676.

   https://doi.org/10.1002/jcp.30011

5. Yan J, Wang J, Zhang H, 2002, An ankyrin repeat-containing protein plays a role in both disease resistance and antioxidation metabolism. *Plant J*, 29(2): 193–202.

   https://doi.org/10.1046/j.0960-7412.2001.01205.x

6. Sekhar KR, Freeman M, 1998, PEST sequences in proteins involved in cyclic nucleotide signalling pathways. *J Recept Signal Transduct Res*, 18(2-3): 113–132.

   https://doi.org/10.3109/10799899809047740

7. Rechsteiner M, Rogers S, 1996, PEST sequences and regulation by proteolysis. *Trends Biochem Sci*, 21(7): 267–271.

   https://doi.org/10.1016/s0968-0004(96)10031-1

8. Salazar-Retana AL, Maruri-López I, Hernández-Sánchez IE, *et al.*, 2019, PEST sequences from a cactus dehydrin regulate its proteolytic degradation. *PeerJ*, 7: e6810.

   https://doi.org/10.7717/peerj.6810

9. Mori T, Li Y, Hata H, *et al.*, 2002, NIRF, a novel RING finger protein, is involved in cell-cycle regulation. *Biochem Biophys Res Commun*, 296(3): 530–536.

   https://doi.org/10.1016/s0006-291x(02)00890-2

10. Wang DY, Hong Y, Chen YG, *et al.*, 2019, PEST-containing nuclear protein regulates cell proliferation, migration, and invasion in lung adenocarcinoma. *Oncogenesis*, 8(3): 1–14.

11. Wu DD, Gao YR, Li T, *et al.*, 2018, PEST-containing nuclear protein mediates the proliferation, migration, and invasion of human neuroblastoma cells through MAPK and PI3K/AKT/mTOR signaling pathways. *BMC Cancer*, 18(1): 1–15.

    https://doi.org/10.1186/s12885-018-4391-9

12. Dong P, Fu H, Chen L, *et al.*, 2020, PCNP promotes ovarian cancer progression by accelerating β-catenin nuclear accumulation and triggering EMT transition. *J Cell Mol Med*, 24(14): 8221–8235.

    https://doi.org/10.1111/jcmm.15491

13. PDB Consortium, 2019, Protein data bank: The single global archive for 3D macromolecular structure data. *J Nucleic Acids Res*, 47(D1): D520–D528.

14. Wang J, Lisanza S, Juergens D, *et al.*, 2021, Deep learning methods for designing proteins scaffolding functional sites. *bioRxiv*, 2021: 468128.

    https://doi.org/10.1101/2021.11.10.468128

15. Jumper J, Evans R, Pritzel A, *et al.*, 2021, Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873): 583–589.

16. Roy A, Kucukural A, Zhang Y, 2010, I-TASSER: A unified platform for automated protein structure and function prediction. *Nat Protoc*, 5(4): 725–738.

    https://doi.org/10.1038/nprot.2010.5

17. Kim DE, Chivian D, Baker D, 2004, Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res*, 32(Suppl 2): W526–W531.

    https://doi.org/10.1093/nar/gkh468

18. Clementi C, 2021, Fast track to structural biology. *Nat Chem*, 13(11): 1032–1034.

    https://doi.org/10.1038/s41557-021-00814-y

19. Pennisi E, 2021, Protein structure prediction now easier, faster. *Science*, 373(6552): 262–263.

https://doi.org/10.1126/science.373.6552.262

20. Yang Z, Lasker K, Schneidman-Duhovny D, *et al*., 2012, UCSF chimera, MODELLER, and IMP: An integrated modeling system. *J Struct Biol*, 179(3): 269–278.

https://doi.org/10.1016/j.jsb.2011.09.006

21. Lengths M, Angles M, 2018, Limitations of structure evaluation tools errat. *Quick Guidel Comput Drug Des*, 16: 75.

22. Laskowski R, MacArthur M, Thornton J, 2006, PROCHECK: Validation of Protein-Structure Coordinates. Hoboken, New Jersey: Wiley.

23. Benkert P, Künzli M, Schwede T, 2009, QMEAN server for protein model quality estimation. *Nucleic Acids Res*, 37(Suppl 2): W510–W514.

https://doi.org/10.1093/nar/gkp322

24. Kutzner C, Páll S, Fechner M, *et al*., 2019, More bang for your buck: Improved use of GPU nodes for GROMACS 2018. *J Comput Chem*, 40(27): 2418–2431.

https://doi.org/10.1002/jcc.26011

25. Hollingsworth SA, Dror R, 2018, Molecular dynamics simulation for all. *Neuron*, 99(6): 1129–1143.

https://doi.org/10.1016/j.neuron.2018.08.011

26. ProtParam E, 2017, ExPASy-ProtParam Tool. ProtParam E.

https://doi.org/10.7717/peerj.10143/table-2

27. Gasteiger E, Hoogland C, Gattiker A, *et al*., 2005, Protein Identification and Analysis Tools on the ExPASy Server. Berlin: Springer. p571-607.

https://doi.org/10.1385/1-59259-890-0:571

28. Suhaibun, S.R., Elengoe A, Poddar R, 2020, Technology advance in drug design using computational biology tool. *Malays J Med Health Sci*, 16(110): 2636–9346.

29. Geourjon C, Deleage G, 1995, SOPMA: Significant improvement in protein secondary structure prediction by c prediction from alignments and joint prediction. *Comput Appl Biosci*, 11(6): 681–684.

https://doi.org/10.1093/bioinformatics/11.6.681

30. Singh N, Upadhyay S, Jaiswar A, *et al*., 2016, In silico analysis of protein. *JSM Bioinform Genom Proteom*, 1(2): 1007.

31. Warde-Farley D, Donaldson SL, Comes O, *et al*., 2010, The GeneMANIA prediction server: Biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res*, 38(Suppl 2): W214–W220.

https://doi.org/10.1093/nar/gkq537

32. Szklarczyk D, Gable AL, Nastou KC, *et al*., 2021, The STRING database in 2021: customizable protein protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res*, 49(D1): D605–D612.

https://doi.org/10.1093/nar/gkaa1074

33. Shannon P, Markiel A, Ozier O, *et al*., 2003, Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res*, 13(11): 2498–2504.

https://doi.org/10.1101/gr.1239303

34. Zhou G, Soufan O, Ewald J, *et al*., 2019, NetworkAnalyst 3.0: A visual analytics platform for comprehensive gene expression profiling and meta-analysis. *Nucleic Acids Res*, 47(W1): W234–W241.

https://doi.org/10.1093/nar/gkz240

35. Kuleshov MV, Jones MR, Rouillard AD, *et al*., 2016, Enrichr: A comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res*, 44(W1): W90–W97.

https://doi.org/10.1093/nar/gkw377

36. Chan JN, Nislow C, Emili A, 2010, Recent advances and method development for drug target identification. *Trends Pharmacol Sci*, 31(2): 82–88.

https://doi.org/10.1016/j.tips.2009.11.002

37. Huang DW, Sherman BT, Lempicki R, 2009, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*, 4(1): 44–57.

https://doi.org/10.1038/nprot.2008.211

38. Kumar S, Stecher G, Li M, *et al*., 2018, MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol*, 35(6): 1547–1549.

https://doi.org/10.1093/molbev/msy096

39. Muhire BM, Varsani A, Martin D, 2014, SDT: A virus classification tool based on pairwise sequence alignment and identity calculation. *PLoS One*, 9(9): e108277.

https://doi.org/10.1371/journal.pone.0108277

40. Gronau I, Moran S, 2007, Optimal implementations of UPGMA and other common clustering algorithms. *Inform Process Lett*, 104(6): 205–210.

https://doi.org/10.1016/j.ipl.2007.07.002

41. Hua GJ, Hung CL, Lin CY, *et al*., 2017, MGUPGMA: A fast UPGMA algorithm with multiple graphics processing units using NCCL. *Evol Bioinform Online*, 13: 1176934317734220.

https://doi.org/10.1177/1176934317734220

42. Ugwu SO, Apte S, 2004, The effect of buffers on protein conformational stability. *Pharm Technol*, 28(3): 86–109.

43. Ikai A, 1980, Thermostability and aliphatic index of globular proteins. *J Biochem*, 88(6): 1895–1898.

44. Yochum GS, Ayer DE, 2002, Role for the mortality factors MORF4, MRGX, and MRG15 in transcriptional repression via associations with Pf1, mSin3A, and transducin-like enhancer of split. *Mol Cell Biol*, 22(22): 7868–7876.

https://doi.org/10.1128/mcb.22.22.7868-7876.2002

45. Durand S, Bruelle M, Bourdelais F, *et al*., 2021, RSL24D1 sustains steady-state ribosome biogenesis and pluripotency translational programs in embryonic stem cells. *bioRxiv*, 2021: 443845.

https://doi.org/10.1101/2021.05.27.443845

46. Doublié S, Kapp U, Aberg A, *et al*., 1996, Crystallization and preliminary X-ray analysis of the 9 kDa protein of the mouse signal recognition particle and the selenomethionyl-SRP9. *FEBS Lett*, 384(3): 219–221.

https://doi.org/10.1016/0014-5793(96)00316-x

47. Zhang Y, Cao X, Li P, *et al*., 2020, PSMC6 promotes osteoblast apoptosis through inhibiting PI3K/AKT signaling pathway activation in ovariectomy induced osteoporosis mouse model. *J Cell Physiol*, 235(7-8): 5511–5524.

https://doi.org/10.1002/jcp.29261

48. Xu CR, Lee S, Ho C, *et al*., 2009, Bmi1 functions as an oncogene independent of Ink4A/Arf repression in hepatic carcinogenesis. *Mol Cancer Res*, 7(12): 1937–1945.

https://doi.org/10.1158/1541-7786.mcr-09-0333

49. Khan NH, Chen HJ, Fan Y, *et al*., 2022, Biology of PEST-containing nuclear protein: A potential molecular target for cancer research. *Front Oncol*, 12: 784597.

50. Makarava N, Chang JC, Molesworth K, *et al*., 2020, Posttranslational modifications define course of prion strain adaptation and disease phenotype. *J Clin Investig*, 130: 4382–4395.