Bulletin of Environment, Pharmacology and Life Sciences
Bull. Env. Pharmacol. Life Sci., Vol 8 [9] August 2019 : 49-55
©2019 Academy for Environment and Life Sciences, India
Online ISSN 2277-1808
Journal's URL:http://www.bepls.com
CODEN: BEPLAD
Global Impact Factor 0.876
Universal Impact Factor 0.9804
NAAS Rating 4.95

ORIGINAL ARTICLE **OPEN ACCESS**

Hypothetical Protein Aeg74552.1 of *Phoenix dactylifera* can used as Anti Hepatocellular Carcinoma Agent

Saad Ur Rehman, Hussain Ali, Anum Munir, Azhar Mehmood, Sajid Khan, Muhammad Rizwan,
Dept. of Bioinformatics, Govt. Postgraduate College, Mandian, Abbottabad
1-Department of Bioinformatics, Govt. Post Graduate College Mandian Abbottabad, Pakistan
Email: saadre23@gmal.com

ABSTRACT
Phoenix dactylifera, commonly known as date or date palm, belongs to genus Phoenix and family Arecaceae, and its fruit posseses high nutritional and therapeutic value with significant anticancer, antifungal, antibacterial, antioxidant properties, immunostimulant activities and provide potential health benefits to the consumer. Hypothetical proteins of Phoenix dactylifera are present in plant which function and structure is not predicted. In this study functional and structural analysis of the HP AEG74552.1 of Phoenix dactylifera is conducted using Bioinformatics databases and tools. The homology modeling of the AEG74552.1 is done with SWISS-MODEL server showing less identity with templates, which manifest that the protein structure is not predicted. So the Ab-initio method was conducted to generate a 3D structure. Validations parameters and quality evaluations determine the stability of the generated protein structure, which has the genuinely good quality. As Phoenix dactylifera has the anticancer properties so we perform docking of this protein with the hepatocellular proteins which show positive results. In future recommendations, this protein would help to expect the structure and functions of several other HP in different species. The comparison studies of this work show that this protein can used as a model against its similar proteins.
Keywords: *Hypothetical Proteins, homology modeling, Hepatocellular Carcinoma, Docking, Phoenix dactylifera.*

Received 11.05.2019 Revised 23.06.2019 Accepted 15.07. 2019

INTRODUCTION

Hypothetical protein HP is a protein which is predicted from ORFs without any experimentally translation validation. The many hypothetical proteins are produced from pseudo-genes, but there are some hypothetical proteins which have a very high possibility of being expressed. These protein structures may be from half of the coding regions of potential proteins in a genome [1]. The function of hypothetical proteins has not been described and no significant information is present. HPs cover have the big ratio of the protein coding regions in several genomes, and in each newly sequenced genome there present almost 20–40% of hypothetical proteins [2,3]. As the structures and sequences of HPs are still remaining non similar to other different proteins, HPs are also known orphan proteins or proteins of unknown functions [4]. This is challenging task for general biology to understand the cellular functions of hypothetical proteins. There are multiple approaches which annotate function of hypothetical proteins differently [3].

Date palm (*Phoenix dactylifera*) is the most significant fruit crops in the East, Southern Europe, South America, North Africa, India, and Pakistan [5]. *Phoenix dactylifera* show a wide range of diversity for its phenotypic characteristics, like, sugar content, fruit color, flowering time and other fundamental traits [6]. Although the *Phoenix dactylifera* tree is broadly planted in Middle East and North Africa, where it is well known as primary source of food. Every plant part has some nutritional, economic, and medicinal importance. The fruit (date) contains a wide range of nutritionally important components, such as, vitamins, fiber, sugars, and some amounts of proteins and fats [7]. At maturity fruit is high-energy source of food about 72% to 88% sugar content, which is usually sucrose. Date fruit is also a good source of iron, calcium, copper, potassium, chlorine, phosphorus magnesium, and sulfur [5]. *Phoenix dactylifera* fruit also possesses essential nutrients and therapeutic value with major antioxidant, anticancer, antibacterial,

antifungal properties, immunostimulant activities and provide potential health benefits to the consumer. The different varieties of dates, are unique for its medicinal properties as drug [8].

The protein three-dimensional (3D) structure is experimentally generated by nuclear magnetic resonance (NMR) or x-ray crystallography or But these two methods are applied in a high throughput way for structure determination. X-ray crystallography is difficult for proteins to form crystals and that's why it is limited and NMR can be applied to small protein only. Protein function is understood by its structure. Therefore, the 3D structure of a protein molecule directly associated to the biological function of that protein [9]. The classical concept of protein structure formation is folding and the process is initiated by quick development of secondary structural elements, then the actual 3D organization of the tertiary structure fold slowly [10]. Hypothetical proteins Structures are responsible for their biophysical or biochemical functions. 3D structure can be responsible for the function of hypothetical proteins. The protein functions prediction from sequence and structure is a tougher problem, because homologous proteins have often different functions. Most of the methods for the prediction of function depend on identifying similarity in the structure or sequence of hypothetical protein and protein with known function [11]. Computational biochemical and Genetic, tools are usually used for the function prediction of such hypothetical proteins. The 3D structure of hypothetical proteins determined by structural genomics initiatives, and annotate the function by understanding the distinctive features of a protein structure, i.e. mode and nature of functional group or metal ion binding, the fold and regulatory sites similarity with other known proteins functions [3]. Scientists established Numerous computational methods and various tools for the prediction of protein function [12].

Hepatocellular carcinoma (HCC) i.e. Liver cancer is the most significant cancers in the world and also main reason for deaths, as per world health organization (WHO) it is more common in male than females worldwide. According to WHO approximation the 40,710 new cases of HCC and 28,920 mortals because of liver cancers [8, 13]. In Pakistan majority of HCC fluctuates from 3.7%-16% of death causing cancers. The most common reason in Pakistan for HCC is the viral hepatitis B, C and D, and about 87% is caused due to the viral hepatitis C (68%) or hepatitis B (22%) [13].

The current work is based on the prediction of structure and function of a HP protein of Date Palm *Phoenix dactylifera*. Many hypothetical proteins have been found in the proteome of several organisms, and in the Date genome we have found 3 hypothetical protein and this hypothetical protein AEG74552.1 is selected on the basis of homology modeling similarity. Hypothetical protein AEG74552.1, has only the amino acid sequence is available, but structure and function are not present. And after the prediction of the structure and function of HP AEG74552.1 it can act as a ligand molecule for the cure of Hepatocellular carcinoma. We have selected the mutated protein of HCC i.e. P53 and dock it with our predicted hypothetical protein.

MATERIAL AND METHODS

Sequence Retrieval of hypothetical protein:

The sequence of amino acids of this hypothetical protein of date palm *Phoenix dactylifera* AEG74552.1 was downloaded from the NCBI (https://www.ncbi.nlm.nih.gov/).

Physiochemical analysis:

Physiochemical characteristics of hypothetical protein AEG74552.1 were performed utilizing ProtParam tool package. (http://web.expasy.org/protparam/) [14].

Secondary Structure prediction:

The server PSPRIED was used to determine secondary structure prediction (helix, coils and sheets) hypothetical protein[15]. another server SOPMA was used to validate the results from PSPRIED [16].

Subcellular localization:

The SOSUI server predicted the subcellular localization of *phoenix dactylifera*. [17]. The results from the SOSUI server were cross-checked by PSORT [18].

Homology modeling:

The homology modeling for 3D structure of the protein AEG74552.1 was constructed using the total of amino acid sequence of this protein in FASTA format from alignment of known protein structure from SWISS-MODEL [19].

Functional annotation:

The analyzation of function for *Phoenix dactylifera* hypothetical protein AEG74552.1 was performed by two different bioinformatics tools ProtFun 2.2 and ProFunc.

3D Structure:

The 3D structure of this hypothetical protein is predicted using the ITESSER server and cross-validated from PHYRE2 and Protein Structure Prediction Server (PS)^2.

3D Structure Validation and Visualization:

Verify3D and ERRAT checked the errors in predicted 3D structure, and for the structural study and verification SAVES server (http://nihserver.mbi.ucla.edu/SAVES/) is used. The structure visualization of generated protein was performed by PyMol and Discovery studio 2016 [19.20].

Docking of Hepatocellular Carcinoma Protein with Hypothetical Protein:

The docking of two proteins hepatocellular carcinoma protein i.e. P53 selected on the basis of the involvement in cancer with newly predicted hypothetical protein AEG74552.1 was performed by the CLUSPRO server, which is a Protein-Protein docking server. And the results were visualized by the PyMol.

RESULT

Physiochemical characteristics of AEG74552.1:

The physiological characteristics of the amino acids of hypothetical protein AEG74552.1 are observed by ProtParam online tool. The huge majority estimation in this server indicate protein stability and steadiness, in the light of fact that the steadiness and stability is identified with its appropriate function ability. This hypothetical protein was contain to be contain 639 amino acids, molecular weight of 71236.74 Daltons with isoelectric point (PI) of 6.01 which confirmed that the protein is negatively charged. The instability index of this protein was estimated to be 49.14 which confirmed that this hypothetical protein as stable. The GRAVY index of -0.253 is confirmed of a hydrophobicity and solubility of this hypothetical protein. The most abundant amino acid residue was to be Leucine (68) followed by Serine 55 and the least amino acid as Cysteine 8. The sequence had 76 negatively charged amino acids (Asp + Glu) and 64 positively charged amino acids (Arg + Lys). The molecular formula of this protein was found as $C_{3182}H_{4928}N_{874}O_{944}S_{22}$.

Subcellular localization of AEG74552.1:

Subcellular localization of the AEG74552.1 protein which expected that where a protein is survives in a cell. Subcellular localization of this hypothetical protein give information about cellular functions. The information might be applied in understanding disease and drug designing. The subcellular localization of this hypothetical protein was expected to be a cellular role, enzyme class and gene ontology analyzed by SOSUI and cross checked by PSORTb.

Secondary structure of AEG74552.1:

The SOPMA server were used to predict the secondary structure of AEG74552.1 Protein. The random coil was predict to be most predominant (42.88%) followed by Alfa helix that is (36.15%), extended strand (16.90%) and beta turn was (4.07%). The analogous results were predicted form the predict protein and PSIPRED servers. The illustrative of secondary structure of AEG74552.1 is predicted from the PSIPRED server is shown is the figure 2:

Homology modeling of AEG74552.1:

We assume these hypothetical proteins a limitless unexplored field with numerous opportunities, both as industrial as well as medical tools. In-silico analysis may help with determining biological functions of such a hypothetical protein. This can be stimulated by expecting the 3D structure of the query protein. At the point when the experimented structure is approachable, similar and homology modeling can be a useful 3D model to the target protein that is determined with no smaller than one experimentally known protein structure. Homology modeling identified the 3D structure of a query protein sequence build principally with respect to its alignment to other proteins which structure is already known.

To perform homology modeling, the sequence of amino acids of the protein was given as input in SWISS-MODEL server. BLASTP search technique was performed for each protein sequence to identify identical templates for homology modeling. The maximum template identity was 20.65% which demonstrate that AEG74552.1.is novel and no identical template structure is present in any database. The 3D structure of this protein is predicted by Ab-initio method through ITASSER server and cross validate by Phyre2 server which gave 100.0% Confidence in 3D model. This 3D model was analyzed by Discovery Studio 2016 and PyMol as shown in figure 3.

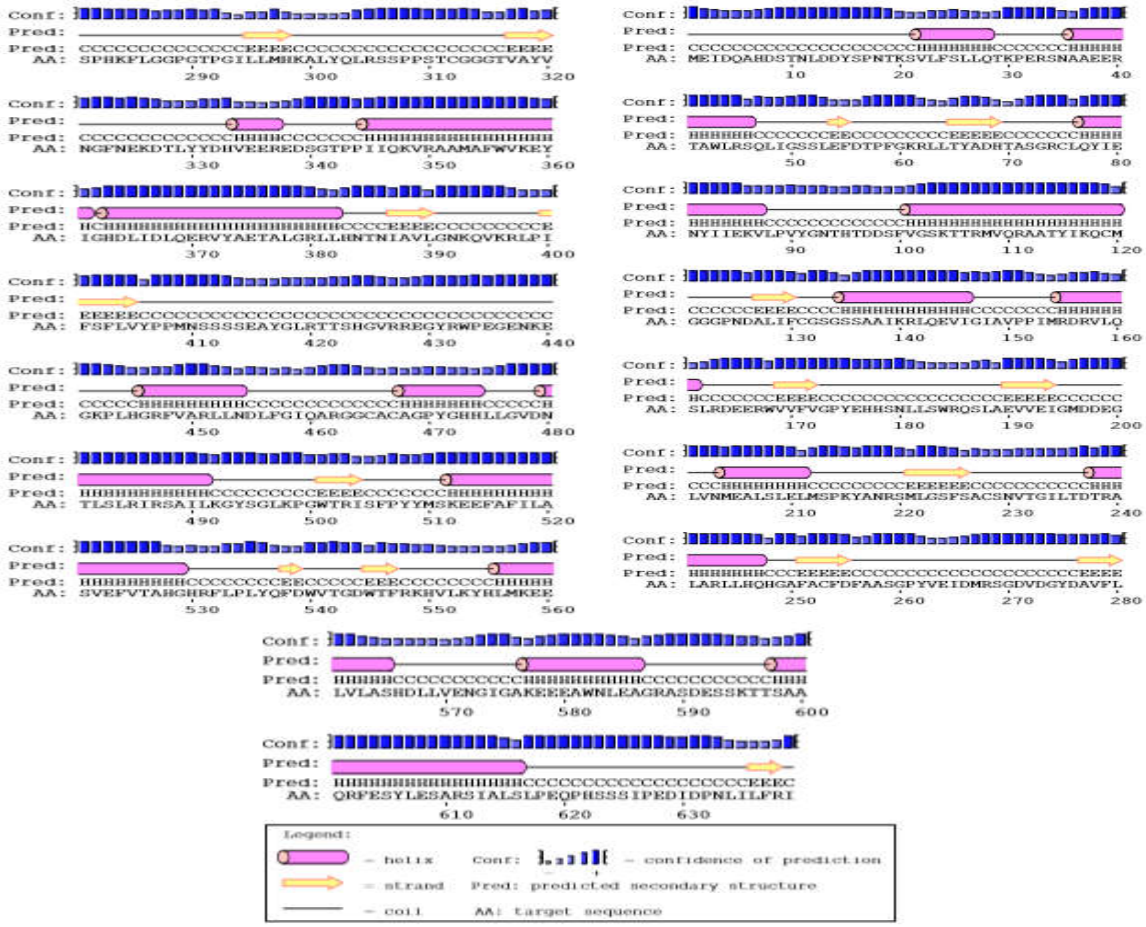

Figure 2: Predicted secondary structure AEG74552.1 by PSIPRED server

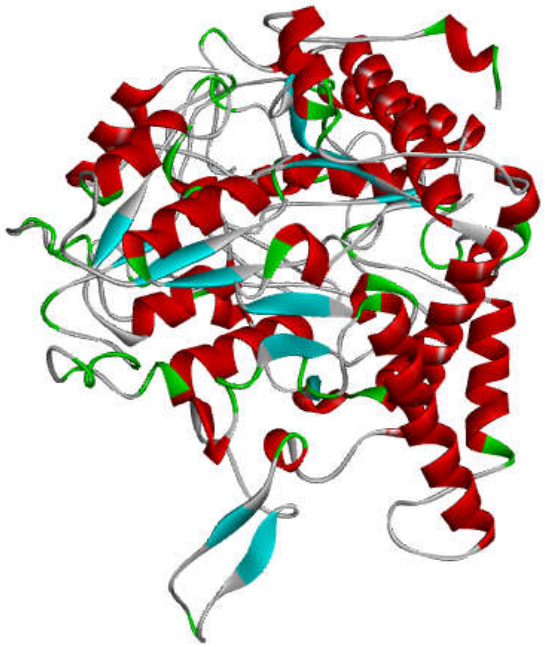

Figure 3: Structural analysis of Date (Phoenix dactylifera) hypothetical protein AEG74552.1

Structure Validation and Visualization:

Reliability and Stability of the build 3D model was checked by ERRAT that examined the statistic calculation of the non-bonded interactions between various atom types, on the basis of characteristic

atomic interactions. The overall quality is identified that is AEG74552.1 Sufficient enough use to this model. As confirmed by Verify3D program, the results indicated that AEG74552.1 of the residues had an average 3D model -1D (amino acid score) score 65.57% meaning that this structures were well-matched and genuinely good.

Functional annotation AEG74552.1:

For functional annotation of AEG74552.1 used two web softwares to search the potential functions. In light of predictions made by ProtFun 2, and ProFunc which show that this protein is involved in cellular role, amino acid biosynthesis, regulatory functions, replication and transcriptions. It was suggested as an immune response, growth factor and metal iron transport as well as action as an enzyme.

Cellular role	Gene ontology	Enzyme class
• **Biosynthesis of cofactors**	• Signal transducer	• Oxidoreductase
• **Cell envelope**	• Metal ion transport	• Transferase
• **Central intermediary metabolism**	• Structural protein	• Ligase
	• Transporter	• Hydrolase
• **Fatty acid metabolism**	• Ion channel	• Lyase
• **Purines and pyrimidines**	• Voltage-gated ion channel	• Isomerase
• **Regulatory functions**	• Cation channel	
• **Transport and binding**	• Transcription	
• **Replication and transcription**	• Stress response	
	• Immune response	
• **Translation**	• Transcription regulation	
• **Cellular processes**	• Receptor Hormone	
• **Energy metabolism**		
• **Amino acid biosynthesis**		

Comparative genome analysis of AEG74552.1

For comparative genomic analysis of AEG74552.1 NCBI BLAST search technique was utilized for comparing genome analysis of AEG74552.1 with other plant genomes. From the result the hypothetical protein AEG74552.1 showed highest identity with other hypothetical proteins of many plants.

Docking of AEG74552.1 with P53:

The docking of the hypothetical protein AEG74552.1 of *Phoenix dactylifera with* the mutated protein of P53 show The interaction between two proteins that they interact to each other and have the binding energy is 6.49 and inhibition constant 8.10, The mutated receptor protein (PDB Id: 6ff9) and the hypothetical protein AEG74552.1 are docked to each other. The good docking complex and the bonding between these proteins shows that this hypothetical protein will be effective to cure the p53 mutated proteins. And this will be the less cost drug protein from Date (*Phoenix dactylifera)* to cure the hepatocellular carcinoma the docking results are shown in the Figure 4.

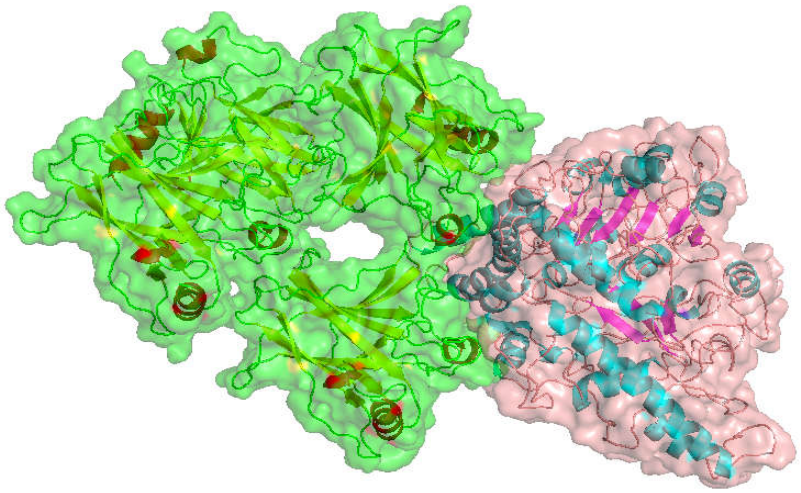

Figure 4: In this figure the mutated receptor protein (green PDB Id: 6ff9) and the hypothetical protein (pink) are docked to each other which show best docked result.

CONCLUSION

Present study focus on the prediction of 3D structure of Hypothetical protein through ab-initio method, and its function on the basis of the 3D structure. The 3D structure is the stable structure as validate from different severs. Its functions are the cellular process, transport, Amino acid biosynthesis. In future recommendations, this protein would help to expect the structure and functions of many other HP in different species. The comparison studies of this work show that this protein can also be used as a model against its similar proteins. Function of the protein shows that the protein is involved in many functions on the basis of similarity with its family proteins. The protein is also utalize as the ligand for the treatment of hepatocellular carcinoma. The purpose of using the date plant protein is that the date fruit shows anti-cancerous properties. The docking of the hypothetical protein of *Phoenix dactylifera* AEG74552.1 with the mutated protein of P53 shows the good docking complex and the bonding between these proteins shows that this hypothetical protein will be effective to cure the p53 mutated proteins. And this will be the less cost drug protein from Date (*Phoenix dactylifera)* to cure the hepatocellular carcinoma in future.

ACKNOWLEDGMENTS

We thank to the Government Post Graduate College Mandian, Abbottabad to provide us the Platform to do this research.

ETHICAL STATEMENT

The material has not been published anywhere. Authors of the manuscript have no financial ties to disclose and have met the ethical adherence.

DISCLOSURE OF INTEREST

The authors declare that they have no competing interests.

DECLARATION OF AUTHORSHIP

All authors have directly participated in the planning, execution, analysis or reporting of this research paper. All authors have read and approved the final version of the manuscript.

CONFLICT OF INTEREST

None.

REFERENCES

1. Enany S. (2014).Structural and functional analysis of hypothetical and conserved proteins of Clostridium tetani. J Infect Public Health. 7(4):296–307.
2. Munir A, Mehmood A, Azam S.(2016). Structural and Function Prediction of Musa acuminata subsp. Malaccensis Protein. 12.
3. Srinivasan B, Kempaiah Nagappa L, Shukla A, Balaram H. (2015). Prediction of substrate specificity and preliminary kinetic characterization of the hypothetical protein PVX_123945 from Plasmodium vivax. Exp Parasitol. 151–152:56–63.
4. Teh B, Choi S, Musa N, Ling F, Cun ST, Salleh A, et al. (2014).Structure to function prediction of hypothetical protein KPN_00953 (Ycbk) from Klebsiella pneumoniae MGH 78578 highlights possible role in cell wall metabolism. BMC Struct Biol.14(1):7.
5. Chao CT, Krueger RR. (2007).The Date Palm (Phoenix dactylifera L.): Overview of Biology, Uses, and Cultivation. HortScience. 42(5):1077–82.
6. Hazzouri KM, Flowers JM, Visser HJ, Khierallah HSM, Rosas U, Pham GM, et al. (2015).Whole genome re-sequencing of date palms yields insights into diversification of a fruit tree crop. Nat Commun. 6(1):8824.
7. Alfaifi SA, Khan MA, Migdadi HM, Jakse J, El-Harty EH, Althamrah MI, et al. Analysis of ESTs from the date palm (Phoenix dactylifera L.) cv. Sukary, an elite variety. :8.
8. Khan F, Khan TJ, Kalamegam G, Pushparaj PN, Chaudhary A, Abuzenadah A, et al. Anti-cancer effects of Ajwa dates (Phoenix dactylifera L.) in diethylnitrosamine induced hepatocellular carcinoma in Wistar rats. BMC Complement Altern Med. 2017 Dec;17(1):418.
9. Toomula N. (2011). Computational Methods for Protein Structure Prediction and Its Application in Drug Design. J Bacteriol Parasitol [Internet]. 2011 [cited 2019 Apr 10];4(12). Available from: https://www.omicsonline.org /computational-methods-for-protein-structure-prediction-and-its-application-in-drug-design-jpb.1000203.pdf
10. Floudas CA. (2007). Computational methods in protein structure prediction. Biotechnol Bioeng. 1;97(2):207–13.
11. Sivashankari S, Shanmughavel P. (2006). Functional annotation of hypothetical proteins – A review. Bioinformation. 2;1(8):335–8.

12. Bharat Siva Varma P, Adimulam YB, Kodukula S. (2015). *In silico* functional annotation of a hypothetical protein from Staphylococcus aureus. J Infect Public Health. 8(6):526–32.
13. Parkash O, Hamid S. (2016). Next big threat for Pakistan Hepatocellular Carcinoma (HCC). J Pak Med Assoc. ;66(6):5.
14. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, et al. (2005). Protein Identification and Analysis Tools on the ExPASy Server. In: Walker JM, editor. The Proteomics Protocols Handbook [Internet]. Totowa, NJ: Humana Press; [cited 2019 Apr 10]. p. 571–607. Available from: http://link.springer.com /10.1385/1-59259-890-0:571
15. Buchan, D. W. A., Minneci, F., Nugent, T. C. O., Bryson, K., & Jones, D. T. (2013). Scalable web services for the PSIPRED Protein Analysis Workbench. Nucleic Acids Research, 41(W1), W349–W357. https://doi.org/10.1093/nar/gkt381
16. Geourjon C, Deléage G. (1995). SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Bioinformatics.11(6):681–4.
17. Yu C-S, Chen Y-C, Lu C-H, Hwang J-K. (2006). Prediction of protein subcellular localization. Proteins Struct Funct Bioinforma. 2;64(3):643–51.
18. Lubec G, Afjehi-Sadat L, Yang J-W, John JPP. (2005).Searching for hypothetical proteins: Theory and practice based upon original data and literature. Prog Neurobiol.;77(1–2):90–127.
19. Arnold K, Bordoli L, Kopp J, Schwede T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics. 15;22(2):195–201.
20. Discovery studio *Dassault Systèmes* BIOVIA, Discovery Studio Modeling Environment, Release (2017), San Diego: Dassault Systèmes, 2016.

CITATION OF THIS ARTICLE

S Ur Rehman, H Ali, A Munir, A Mehmood, S Khan, M Rizwan. Hypothetical Protein Aeg74552.1 of *Phoenix dactylifera* can used as Anti Hepatocellular Carcinoma Agent. Bull. Env. Pharmacol. Life Sci., Vol 8 [9] August 2019: 49-55