International Journal of Zoology and Applied Biosciences
Volume 1, Issue 2, pp: 94-99, 2016

ISSN: 2455-9571

http://www.ijzab.com

Research Article

# AN INSILICO STUDY: CHARACTERIZATION OF WD-REPEAT PROTEIN FAMILY IN DIFFERENT SPECIES OF MOSQUITOS

**Sivakolundu Mahesh Babu, Mathalaimuthu Baranitharan, Shanmugam Dhanasekaran*, Subbaiyan Thushimenan**

Department of Zoology, Annamalai University, Annamalainagar-608 002, Tamilnadu, India.

**Article History:** Received 3rd March 2016; Revised 22nd April 2016; Accepted 23rd April 2016

## ABSTRACT

In the present investigation, the WD-repeat (WDR) proteins comprise an astonishingly diverse superfamily of regulatory proteins. To date, genome-wide characterization of this family has only been conducted in mosquito and little is known about WDR genes in mosquito (*Aedes*, *Anopheles* and *Culex*). This study identified 15 mosquito WDR genes in the latest genome and the WDR family contained a smaller number of identified genes compared to different species of mosquitoes. The WDR proteins were identified and classified in to different subfamilies based on their distinct domain organizations. Although many characteristics of the protein family are similar to those species, several features are quite distinct. Our result of Insilco analysis indicated the existence of well-conserved subfamilies. Moreover, comparative genomic analysis showed that the gene structures of the WDR protein were highly conserved across some different lineage species. Through functional divergence analysis, a substantial divergence was found between WDR protein subfamilies.

**Keywords:** WD-repeat, *Aedes aegypti*, Physiochemical Properties, Sequence alignment, Phylogenetic tree.

## INTRODUCTION

The so-called WD-repeat (WDR) proteins comprise an astonishingly diverse superfamily of regulatory proteins, representing the breath of biochemical mechanisms and cellular processes (Steven *et al.,* 2003). They are found in all eukaryotes but usually not in prokaryotes (Neer *et al.*, 1994). Members of WDR proteins family do not have an immediately obvious common function but are involved and have been found to play key roles in diverse cellular functions and pathways. A recognized general role of WDRs is to mediate protein-protein interactions and coordinate multi-protein complex formations in many events such as cell division, cell-fate determination, signal transduction, transmembrane signaling, premRNA splicing and mRNA modification, nuclear export, cytoskeletal assembly and dynamics, vesicle fusion and vesicular traffic, protein trafficking, apoptosis, and are especially prevalent in chromatin modification and transcriptional mechanisms (Thompson *et al.,* 1994). The underlying common feature of all the members of WDR family is their coordination of multi-protein complex assemblies. These repeats provide a stable platform or scaffold on which large protein or protein-DNA complexes can assemble. As determined by the available WD40 domain complex structures, the WDR proteins have three distinct surfaces (top, bottom and circumference) that can be exploited for interaction with other proteins (Chen *et al.,* 2011). When present in a protein, the WD motif is typically found as several tandem repeat units. Since the first WD40 domain structure was determined, tens of WD40 structures have been determined to date (Xu and Min, 2011), including a mammalian Gβ subunit of heterotrimeric GTPases, repeated WD units that form a series of four-stranded, antiparallel beta sheets, which fold into a higher-order structure termed as β-propeller. This structure can be visualized as a short, open cylinder where the strands form the walls (Smith *et al.,* 1999). Peculiar features of the WDR gene family are- i) low sequence conservation although high evolutionary conservation; ii) co-occurrence of the WD40 domain with other domains; iii) interaction with multiple proteins to form large complexes and; iv) the functional diversity. Despite of the global reduction in malaria mortality by 42% from 2000 to 2012, malaria continues to be a major public health problem threatening 3.4 billion people across 97 countries and causing deaths of ~627,000 people in 2012 (WHO, 2013).

*Corresponding author: Dr. S. Dhanasekaran, Assistant Professor, Department of Zoology, Annamalai University, Annamalainagar - 608 002, Tamil Nadu, India, Email: mahenithi@gmail.com, Mobile: +91 9443260002.

## MATERIAL AND METHODS

### Identification of WDR protein in mosquito

This study used amino acid sequences of all known WDR proteins that were identified by using the key term "WD-repeat protein in mosquito" in protein knowledge database (UniprotKB) available at http://www.uniprot.org/. a collection of 15 sequence entries was identified from different mosquito species(Q7QF60, Q1HQN6, A0A023EIB1, A0A023ET56, A0A023ENL4, W5J9I0, W5J9Z7, A0A023EPG3, T1DKR9, W5JSZ9, Q1HRQ2, Q7PP77, C7BB02, C7BB04, C7BB04)genes. The different mosquito families included in this study (*Anopheles gambiae, Aedes aegypti, Aedes albopictus, Anopheles darling, Anopheles aquasalis, Anopheles quadriannulatus, Anopheles merus*)

### Physiochemical Parameters Generation Using Various Online Tools

Physiochemical data were generated from the ProtParam software using ExPASy server (Gough *et al.,* 2001). The proteomic server of Swiss Institute of Bioinformatics. FASTA sequence format were applied for subsequent analysis. Different tools in the Proteomic server (ProtParam and Compute pI /Mw) were applied to figure out different physiochemical properties of WDR protein sequences in mosquito. The amino acids number, molecular weight (kilodalton) and pI values were deduced by using compute pI/Mw, and the atomic compositions, values of instability index, aliphatic index and grand average of hydropathicity (GRAVY) of WDR protein sequences in mosquito were derived using the ProtParam tool, available at ExPASy.

### Secondary structure analysis:

Secondary structure analysis was carried out using SOPMA server https://npsa-prabi.ibcp.fr/cgi-bin/secpred_sopma.pl (Kyte and Doolittle, 1982), and SOSUI server http://harrier.nagahama-i-bio.ac.jp/sosui/sosui_submit.html (Combet *et al.,* 2000).

### Multiple sequence alignment and phylogenetic analysis:

Multiple sequence alignment was performed using the Clustal W program (http://www.ebi.ac.uk/clustalw/) available at the European Bioinformatics Institute web site (Thompson *et al.,* 1994). To investigate the evolutionary relationship between the putative WDR protein sequences identified here and other sequence alignment represented by CLC-Bio sequence viewer http://www.clcbio.com/index.php?id=28 (CLC Sequence Viewer) used for construction by Neighbor-joining (NJ) method using 100 bootstrap values (Larkin *et al.,* 2007). Protein sequences used for phylogenetic analysis were extracted from protein knowledge database (UniprotKB). Amino acid sequences used for analysis.

### 3D Structure Prediction Using Homology Modeling Approach

Protein 3D structure is very important in understanding the protein interactions, functions and their localization. Homology modeling is the most common structure prediction method. Homology modeling was done using workspace. The SWISS-MODEL Workspace is a web-based integrated service dedicated to protein structure homology modeling (Marco *et al.,* 2014). It assists and guides in building protein homology models at different levels of complexity. Retrieve all the 15 sequence entries from Swiss Prot:www.expasy.org/sprot and carry out homology modelling using SWISS MODEL: http://swissmodel.expasy.org/ (Arnold *et al.,* 2006). Template identification is done for each sequence using the template identification tool provided in SWISS MODEL. Carry out the modeling using the same server in alignment mode. Follow the online instructions to carry out the task and get the results. In order to facilitate the use of alignments in different formats, the submission is implemented as a three step procedure: Prepare a multiple sequence alignment We used CLUSTAL W to carry out multiple sequence alignment of the query protein sequences with the homologous target sequences obtained using the "template Idendification" tool in SWISS-MODEL. The multiple sequence alignment was carried out using the site http://www.ebi.ac.uk/Tools/msa/clustalw2/. Submit your alignment to the Workspace Alignment Mode--Possible formats are: FASTA, MSF, CLUSTALW, PFAM and SELEX.One may either upload the file or cut and paste. One should not forget to specify the correct alignment format. The sequence of the template structure was selected validation for generated models was done (Benkert *et al.,* 2011). Quality and reliability of structure was checked by several structure assessment methods including Z-score and Ramachandram plots.This RAMPAGE tool was used to determine the Ramachandran plot to assure the quality of the model. The result of the Ramachandran plot showed 98.0% of residues in favorable region representing that it is a reliable and good quality model. A model having more than 90% residues in favorable region is considered as good quality model (Lovell *et al.,* 2002).

## RESULTS AND DISCUSSION

In this study primary structure of The 15 WD repeat protein sequences were retrieved from Expasy's Prot Param server (http://expasy.org/cgi-bin/protparam) using the gene sequence and the results are shown in Table 1. Results showed that WD repeat protein had amino acid residues and the estimated molecular weight (98560.67). The maximum number of amino acid present in the sequence was found. The calculated isoelectric point (pI) is useful for at pI the solubility is least and the mobility in an electric field is zero. Isoelectric point (pI) is the pH at which the surface of protein is covered with charge but net charge of protein is zero. The calculated isoelectric point (pI) was computed to be 8.03. The computed value is more than 7 indicate that the protein is basic. The Grand Average Hydropathicity (GRAVY) value is low -0.113, indicates

better interaction of the protein with water. The secondary structure is composed of alpha helix and beta sheets and the secondary structure is predicted using SOPMA and SOSUI. Table 2 presents the comparative analysis of SOSUI and SOPMA from which it is clear that random coil is predominantly present when the structure was predicted both by SOPMA and SOSUI, followed by extended strand and alpha helix. The secondary structure prediction was done and random coil was found to be frequent (31.53%) followed by Extended strand (42.36%) and alpha helix was found to be least frequent (8.05%).A total of 15 WD repeat protein sequences from different source organisms subjected to phylogenetic tree construction revealed major clusters shown in (Figure 1). The multiple accessions of WD repeat protein sequences were placed closely in the clusters signifying the greater degree of sequence level similarity. Similar phylogenetic tree revealing clustering of WD repeat protein sequences based on different source organism has been reported. The tertiary structure was modelled by Swiss model workspace by using the templates from PDBSum (Figure 2) the modelled structure showed Number of residues in favoured region 562 ( 94.5%) Number of residues in allowed region 27 (4.5%) Number of residues in outlier region 6(1.0%). The modelled structure was validated by Rampage server**,** Ramachandran plot was plotted (Figure 3).

**Table 1.** Retrieved sequences, source, species name and their accession number physicochemical characterization of Wd repeat proteins.

| S. No. | Accession No. | Source Organisms | Number of amino acids | Molecular weights | Theoretical pi | Gravy |
|---|---|---|---|---|---|---|
| 1 | Q7QF60 | *Anopheles gambiae* | 934 | 98560.67 | 7.83 | -0.339 |
| 2 | Q1HQN6 | *Aedes aegypti* | 331 | 37020.68 | 5.19 | -0.209 |
| 3 | A0A023EIB1 | *Aedes albopictus* | 179 | 19806.44 | 5.28 | -0.155 |
| 4 | A0A023ET56 | *Aedes albopictus* | 425 | 46011.63 | 6.00 | -0.274 |
| 5 | A0A023ENL4 | *Aedes albopictus* | 252 | 28054.87 | 9.01 | -0.305 |
| 6 | W5J9I0 | *Anopheles darlingi* | 351 | 38993.79 | 6.67 | -0.387 |
| 7 | W5J9Z7 | *Anopheles darlingi* | 314 | 35046.82 | 6.44 | -0.235 |
| 8 | A0A023EPG3 | *Aedes albopictus* | 359 | 39950.14 | 7.90 | -0.440 |
| 9 | T1DKR9 | *Anopheles aquasalis* | 456 | 49663.80 | 5.78 | -0.254 |
| 10 | W5JSZ9 | *Anopheles darlingi* | 302 | 32890.11 | 5.37 | -0.150 |
| 11 | Q1HRQ2 | *Aedes aegypti* | 311 | 34891.50 | 8.03 | -0.336 |
| 12 | Q7PP77 | *Anopheles gambiae* | 327 | 36194.59 | 5.26 | -0.272 |
| 13 | C7BB02 | *Anopheles arabiensis* | 174 | 19042.22 | 4.67 | -0.113 |
| 14 | C7BB04 | *Anopheles quadriannulatus* | 174 | 19042.22 | 4.67 | -0.113 |
| 15 | C7BB04 | *Anopheles merus* | 174 | 19035.17 | 4.77 | -0.130 |

Table 2. Secondary structure information of WD repeat proteins.

| S. No. | Accession No. | Source Organisms | Alpha Helix | Extended Strand | Beta Turn | Random Coil | Average Hydrophobicity |
|---|---|---|---|---|---|---|---|
| 1 | Q7QF60 | *Anopheles gambiae* | 20.56% | 24.20% | 9.96% | 45.29% | -0.338758 |
| 2 | Q1HQN6 | *Aedes aegypti* | 8.76% | 41.99% | 13.60% | 35.65% | -0.209366 |
| 3 | A0A023EIB1 | *Aedes albopictus* | 24.02% | 29.05% | 11.73% | 35.20% | -0.155307 |
| 4 | A0A023ET56 | *Aedes albopictus* | 4.47% | 42.82% | 14.82% | 37.88% | -0.274118 |
| 5 | A0A023ENL4 | *Aedes albopictus* | 17.46% | 39.68% | 11.51% | 31.35% | -0.304762 |
| 6 | W5J9I0 | *Anopheles darlingi* | 8.83% | 42.45% | 14.65% | 31.53% | -0.386895 |
| 7 | W5J9Z7 | *Anopheles darlingi* | 11.46% | 42.36% | 14.65% | 31.53% | -0.235350 |
| 8 | A0A023EPG3 | *Aedes albopictus* | 9.47% | 36.77% | 12.81% | 40.95% | -0.440111 |
| 9 | T1DKR9 | *Anopheles aquasalis* | 4.39% | 44.74% | 15.13% | 35.75% | -0.253947 |
| 10 | W5JSZ9 | *Anopheles darlingi* | 19.21% | 35.76% | 16.23% | 28.81% | -0.150331 |
| 11 | Q1HRQ2 | *Aedes aegypti* | 12.22% | 43.73% | 16.40% | 27.65% | -0.336013 |
| 12 | Q7PP77 | *Anopheles gambiae* | 16.82% | 36.70% | 9.17% | 37.31% | -0.272477 |
| 13 | C7BB02 | *Anopheles arabiensis* | 8.05% | 46.55% | 14.37% | 31.03% | -0.112644 |
| 14 | C7BB04 | *Anopheles quadriannulatus* | 8.05% | 46.55% | 14.37% | 31.03% | -0.112644 |
| 15 | C7BB04 | *Anopheles merus* | 6.32% | 45.98% | 16.09% | 31.61% | -0.129885 |

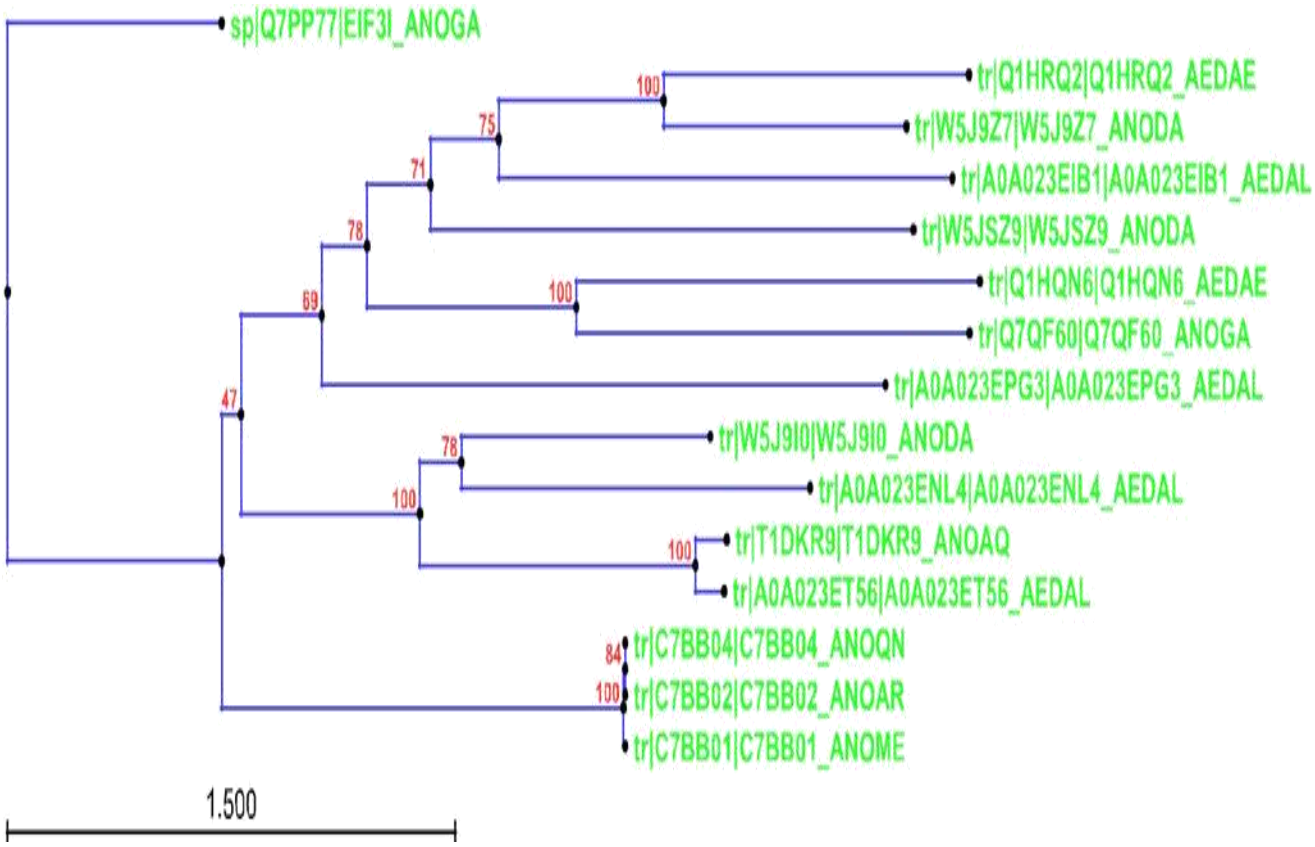


**Figure 1.** Phylogenetic analysis of WD-repeat protein sequences from different species of mosquitoes using Neighbor-joining (NJ) method.

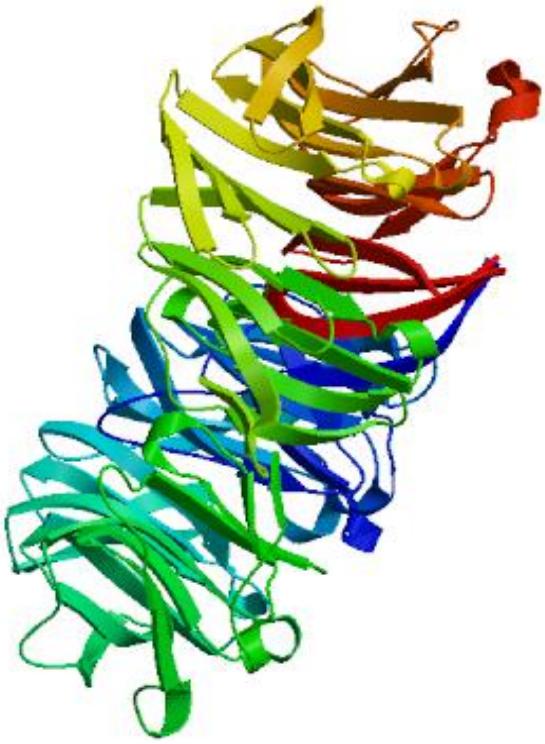

**Figure 2.** Swiss-model using predicted structure of 1pev.1.A wd repeat proteins.

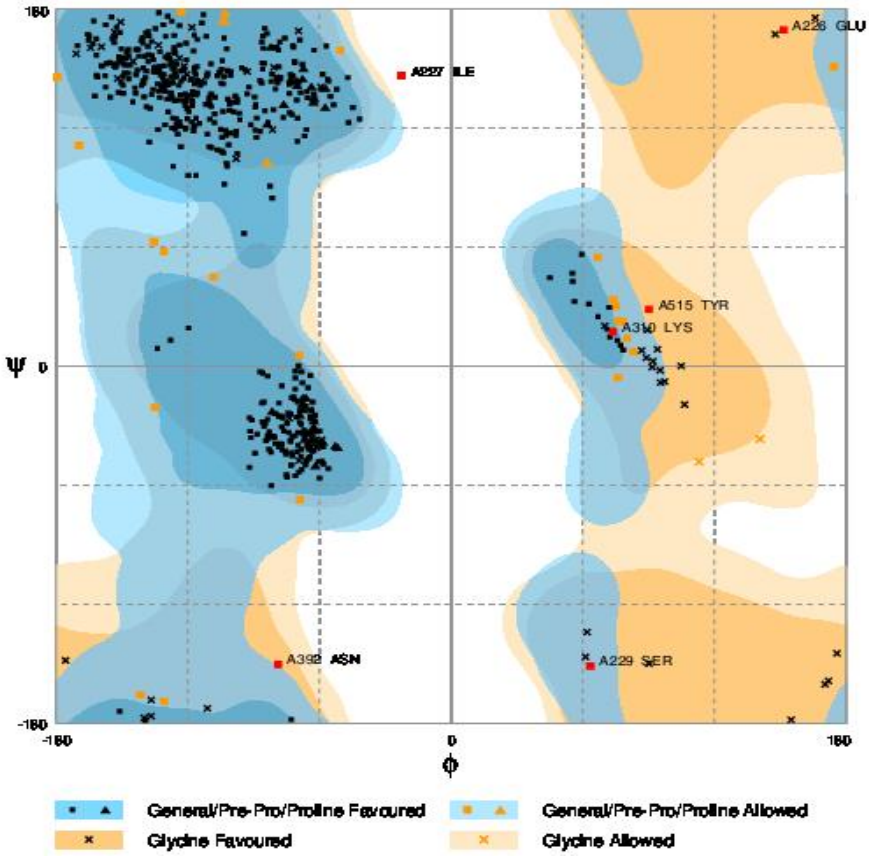


**Figure 3.** Validation of modelled structure of WD repeat protein using rampage server [Number of residues in favoured region (~98.0% expected): 562 (94.5%). Number of residues in allowed region (~2.0% expected): 27 (4.5%). Number of residues in outlier region: 6 (1.0%)].

## CONCLUSION

Mosquito borne diseases gained dominant position by life threatening hazards. Human population are altering and polluting the environment and encouraging vectors which subsequently causes diseases. In this study WD repeat protein were selected Physicochemical characterization were performed by computing theoretical isoelectric point (pI), molecular weight, total number of positive and negative residues, extinction coefficient, instability index, aliphatic index and grand average hydropathy (GRAVY). Functional analysis of these proteins was performed by SOSUI server. For these proteins disulphide linkages, motifs and profiles were predicted. Secondary structure analysis revealed that random coils dominated among secondary structure elements followed by alpha helix, extended strand and beta turns for all sequences. The modelling of the three dimensional structure of the proteins were performed by homology program Swiss model The models were validated using protein structure checking tools. The sequences were characterized for homology search, multiple sequences alignment, biochemical features, phylogenetic tree construction search using various bioinformatics tools.

## ACKNOWLEDGEMENTS

The authors are thankful to the professor and Head, Department of Zoology Annamalai University, Chidambaram, Tamilnadu, India.

## REFERENCES

Arnold, K., Bordoli, L., Kopp, J., Schwede, T., 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics.,* 22, 195-201.

Benkert, P., Biasini, M., and Schwede, T., 2011. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics.,* 27, 343-350.

Chen, C.K., Chan, N.L., Wang, A.H., 2011. The many blades of the beta-propeller proteins: Conserved but versatile. *Trends. Biochem. Sci.,* 36, 553-561.

CLC, 2012. Sequence Viewer. http://www.clcbio.com/index.php?id=28 (accessed on 27 January).

Combet, C., Blanchet, C., Geourjon, C., Deléage, G., 2000. NPS@: Network Protein Sequence Analysis. *TIBS*, 291, 147-150.

Gough, J., Karplus, K., Hughey, R., Chothia, C., 2001. Superfamily: HMMs representing all proteins of known structure. SCOP sequence searches, alignments and genome assignments *J. Mol. Biol.,* 313, 903-919.

Kyte, J. and Doolittle, R.F., 1982. A simple method for displaying the hydropathic character of a protein. *J. Mol. Bio.,* 157, 105.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., William, H.M., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J. and Higgins, D.G., 2007. Clustal W and Clustal X Version 2.0. *Bioinformatics*, 23, 2947-2948.

Lovell, S.C., Davis, I.W., Arendall III, W.B., de Bakker, P.I.W., Word, J.M., Prisant, M.G., Richardson, J.S., and Richardson, D.C., 2002. Structure validation by Calpha geometry: phi, psi and Cbeta deviation. Proteins: Structure, Function and Genetics, 50, 437-450.

Marco, Biasini., Stefan, Bienert., Andrew, Waterhouse., Konstantin, Arnold., Gabriel, Studer., Tobias, Schmidt., Florian, Kiefer., Tiziano, Gallo, Cassarino., Martino, Bertoni., Lorenza, Bordoli., Torsten, Schwede., 2014. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic. Acids. Res.,* 42(W1): W252-W258. doi: 10.1093/nar/gku340.

Neer, E.J., Schmidt, C.J., Nambudripad, R. and Smith, T.F., 1994. The ancient regulatory-protein family of WD-repeat proteins. *Nature,* 371, 297-300.

Smith, T.F., Gaitatzes, C., Saxena, K., Neer, E.J., 1999. The WD repeat: a common architecture for diverse functions. *Trends. Biochem. Sci.,* 24, 181-185.

Steven Van Nocker, S, and Ludwig, P., 2003. The WD-repeat protein superfamily in *Arabidopsis*: conservation and divergence in structure and function. *B.M.C. Genomics.,* 4, 50.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTALW: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic. Aci. Res.,* 22, 4673-4680.

World Health Organization., 2013. World malaria report, Geneva: World Health Organization.

Xu, C., Min, J.R., 2011. Structure and function of WD40 domain proteins. *Protein. Cell.,* 2, 202-214.