

# 'MYB-Related Protein P-Like' is the Candidate Gene Homologus to ZmMYB-IF35 R2R3-Transcription Factor in Wheat (*Triticum aestivum* L.) for Anthocyanin Pathway Modifications

Tayyaba Noor<sup>\*</sup>, Faisal Saeed Awan

Department of Biotechnology, University of Agricultural Science, Faisalabad, Pakistan

## ABSTRACT

Transcription factors belonging to R2R3 family play an essential role in the regulation of secondary metabolites and during stress conditions. Diverse number of such transcription factors has been identified but very few have been characterized in monocots. ZmMYB\_IF35 have been characterized previously in *Zea mays* in regulation of anthocyanin biosynthetic pathway. This study aims at characterizing the R2R3-MYB transcription factors similar to ZmMYB-IF35 in wheat for the purpose of regulation of flavonoid/anthocyanin pathway through computational tools. 'MYB-related protein P-like' (C<sub>1618</sub>H<sub>2565</sub>N<sub>479</sub>O<sub>534</sub>S<sub>15</sub>) is a negatively charged protein, composed of 346 amino acids with molecular weight of 37752.95 (da) and isoelectric point of 5.19. It is highly thermostable and is localized in the nucleus. Gene responsible for its transcription is located on chromosome 4B on LOC119296349 in wheat (*Triticum dicoccoides*). Significant portion of the protein was found to be composed of random coils (50%) and small nonpolar residues. From transmembrane topology predictions we showed that, N-terminal of the protein constitutes of all transmembrane helices. From gene ontology predictions, biological function was showed to be regulation of RNA biosynthesis process majorly and it binds to the nucleic acids/DNA. No phosphorylation was shown as post translational modification apart from glycosylation between 150-200 amino acid residues. Interestingly, PEST sequences were also found within the glycosylation rich region. Ile-58, Gly-64, Leu-77, Glu-103, Glu-34 were shown to be functional residues that can be mutated at the catalytic pocket. The likely centroid ligands that can bind to the binding site 180-184 (residues) and 225-228 (residues) were shown to be dopamine (DA382; DA893) and ruthenium (RA914) with intermediating role in binding to DNA or chromatin. A total of three superfamilies were predicted; PLN03212, REB1 superfamily and SANT superfamily were predicted with SANT MYB-DNA binding superfamily with highest significance near N-terminal of sequence. Our results provide novel insights on the characterization of 'MUB-related protein P-like' in *Triticum dicoccoides* homologous to *Zea mays* ZmMYB-IF35 for the regulation of flavonoid/anthocyanin biosynthesis pathway in wheat.

**Keywords:** Anthocyanins; Photosynthetic; Pollination

## INTRODUCTION

Anthocyanins belong to a class of flavonoids that are involved in imparting characteristic color to the plant tissues. Their production within plant cells is controlled by various developmental factors as well as environmental signals. They are involved in the number of physiological functions such as

pollination, seed dispersal and induction of stress response against low temperature, pathogenic infections, nutrient deficiency and radiations.

Anthocyanins in foliage leaves are not researched enough. Chlorophylls and carotenoids contribute more towards color of foliage than anthocyanins. In leaves of a plant, the anthocyanins

**Correspondence to:** Tayyaba Noor, Department of Biotechnology, University of Agricultural Science, Faisalabad, Pakistan; E-mail: tayyaba-noor@outlook.com

**Received:** 21-Nov-2023, Manuscript No. CSSB-23-28104; **Editor assigned:** 24-Nov-2023, PreQC No. CSSB-23-28104 (PQ); **Reviewed:** 08-Dec-2023, QC No. CSSB-23-28104; **Revised:** 01-Apr-2025, Manuscript No. CSSB-23-28104 (R); **Published:** 09-Apr-2025, DOI: 10.35248/2332-0737.25.13.106

**Citation:** Noor T, Awan FS (2025) 'MYB-Related Protein P-Like' is the Candidate Gene Homologus to ZmMYB-IF35 R2R3-Transcription Factor in Wheat (*Triticum aestivum* L.) for Anthocyanin Pathway Modifications. J Curr Synth Syst Bio. 13:106.

**Copyright:** © 2025 Noor T, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

are believed to be responsible for the red coloration in autumn season. The green color of a leaf is only due to chlorophyll but anthocyanins play their part by letting only yellow green light to pass through and allowing the leaves to excite only to this light, since oxygen radicals are produced by excitation of chlorophyll molecules, thus the oxidative load is reduced in leaves. In foliage, these pigments are located inside the vacuoles and upper as well as lower epidermis thus explaining the diverse locations for anthocyanins.

Gould, in 2004, described in its article, that anthocyanins serve as protectants for photolabile compounds (compounds that become unstable upon exposure to light and are part of defense mechanism in a plant). Moreover, it was also explained that although anthocyanins serve as protectant for photosynthetic components but absorbing excess quantum energy from light and protecting the plant from reducing quantum efficiency, more of the photons than needed are absorbed as a result and hence, deprive chlorophyll b from necessary photons. In short, the photosynthetic efficiency is reduced. However, the gas analysis of red and green leaves of *O. triangularis* indicated that lower CO<sub>2</sub> assimilation happens in red leaves as compared to green leaves but the efficiency of photosystem II remains higher in red leaves as compared to green leaves. Moreover, red leaves were also less photo inhibited than the green ones. In another experiment, the red and green leaves of dogwood plant were exposed to white light for 30 minutes. The results showed 30 percent reduction in quantum efficiency in red leaves and 100 percent reduction in green leaves indicating the importance of anthocyanins for photosystem [1].

These pigments impart red and blue coloration when exposed to acidic and alkaline environment respectively (Figure 1).

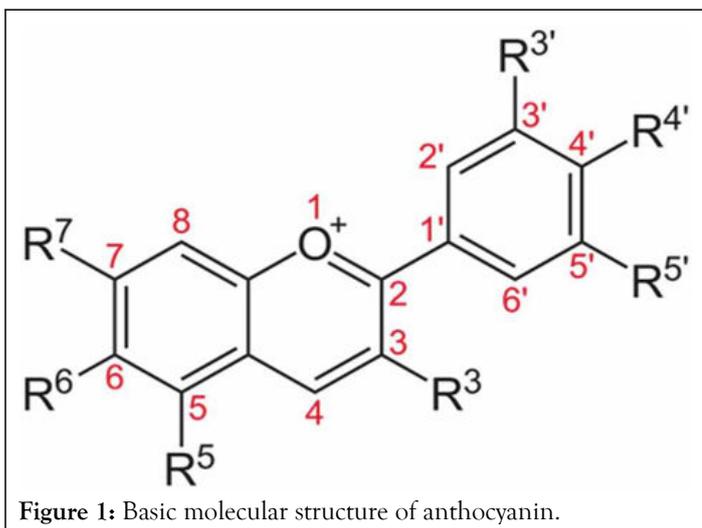


Figure 1: Basic molecular structure of anthocyanin.

The general structure of this molecule is composed of a positive charge on its C-ring hence named as flavylium ion. Empirical formula for a general anthocyanin is C<sub>15</sub>H<sub>11</sub>O<sup>+</sup> with a molecular weight of 207.24724 g/mol. This ion is also termed as 2-phenylchromenylium. These compounds belong to the family of phenolic compounds and are glycoside in nature hence aglycone. They are also called as phytochemicals. Anthocyanins can be grouped into two categories depending upon their nature. They are: Glycosides and acylated.

The glycosylated forms of anthocyanins are also termed as anthocyanins. Cyanidin, delphinidin, pelargonidin, peonidin, malvidin and petunidin are the most common anthocyanidins distributed in the plants. The distribution of these anthocyanidins in fruits and vegetables is 50%, 12%, 12%, 12%, 7% and 7%, respectively.

Expression of anthocyanin biosynthesis genes is highly controlled by MBW protein complex, and R2R3-MYB transcription factors having the central role in the activation of anthocyanin biosynthetic genes [2].

## Transcription factors and MYB

Transcription factors play crucial part in changing or controlling the cellular processes. They can modify the complex traits of plants. The genes of transcription factors contain a number of motifs in them such as; zinc fingers, MYC, bZIP and MYB. These genes are regulated or induced by specific signals such as stress signals. Among these genes, MYB containing transcription factors are large and diverse. MYB belongs to a complex of proteins named as MBW that regulates phenylpropanoid biosynthesis pathway and are involved in epidermal cell fate. Others members of this family are; MYB, bHLH and WDR. The MYB proteins have the ability to bind to DNA thus implicating ABA response or interacting with other transcription factors. V-MYB was the first MYB gene that was identified under the category of oncogene from *Zea mays* (C1). The members of these genes are: c-MYB, A-MYB and B-MYB.

In *Arabidopsis thaliana*, 190 R2R3-MYB genes have been identified. In corn (*Zea mays*), 157 R2R3-MYB encoding genes were identified. In populus, 192 R2R3-MYB encoding genes and 5 3R-MYB genes were found. In soybean, 252 MYBs, including 244 R2R-MYB (2R-MYB) genes, 6 R1R2R3-MYB belonging to 3R-MYB family genes and two R0R1R2R3-MYB belonging to 4R-MYB genes were identified [3].

## R2R3-type MYB transcription factors

R2R3-type MYB transcription factors are mainly found in plants. The structure of this type contains 2 domains; one is N-terminal conserved MYB DNA-binding domain and the other domain is C-terminus modulator region. The C-terminus is responsible for the regulatory activity performed by the protein. The MYB-binding domain contains 4 classes of proteins that are explained in the diagram (Figure 2).

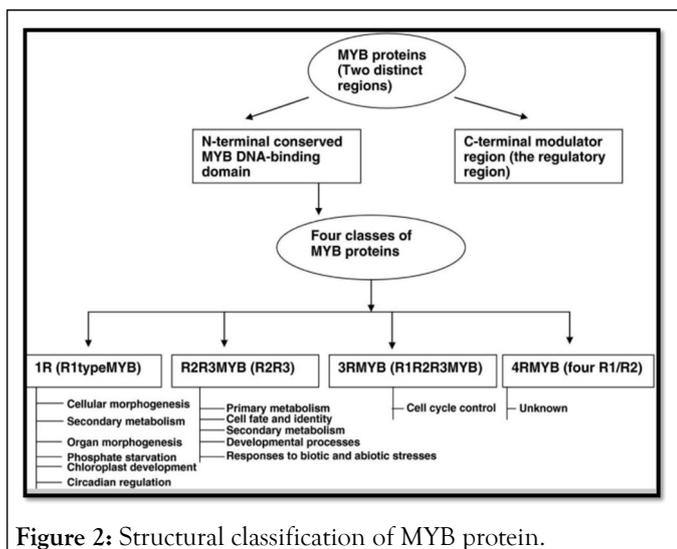


Figure 2: Structural classification of MYB protein.

MYB are essential as they are; essential for the development of plants, they are involved in determining shape of cell and petal morphogenesis, differentiation of cell and its proliferation, involved in the development of trichome and phenylpropanoid metabolism, play role in hormonal responses, help against abiotic stresses and biotic stresses and are involved in regulation of primary and secondary metabolites.

Anthocyanin production in wheat is controlled by many regulatory genes such as *Pc*, *Pan*, *Pg*, *Plb*, *Pp*, *R*, *Ra*, *Rc* controlling purple culm, purple anther, purple glume, purple leaf blade, purple pericarp, red glume, red auricle and red coleoptile respectively. In another study, two loci; *Pp-D1* and *Pp3* on chromosome 7D and 2A having *TaPpm1* (purple pericarp-MYB1) and *TaPpb1* (purple pericarp-bHLH1) genes were identified in regulating anthocyanin biosynthesis in purple pericarps [4].

In *Zea mays*, P1 gene is responsible for the production and regulation of flavonoids (3-deoxy flavonoids) and phlobaphene pigments in floral regions. P1 binds with promoter of *a1* gene and results in the production of anthocyanin biosynthetic pathway enzyme *i.e.*, Dihydroflavonol-4-reductase (DFR). Similarly, in mays, another R2R3-transcription factor; ZmMYB-IF35 controls the accumulation of several distinct phenylpropanoids (through anthocyanin biosynthetic pathway) and phenolic compounds in maize. Herine, et al. showed that ZmMYB-IF35 is capable to activate *a1* promoter through its regions outside the conserved domains and results in regulation of phenylpropanoid biosynthetic pathway. Phenylpropanoid biosynthetic pathway is a part of anthocyanin pathway in plants. This type of transcription factor was also found to be involved in controlling chilling stress in maize suggesting its strong role in the defense mechanism as well as anthocyanin pathway regulation. Therefore, in the current study, we aimed at finding a homologous protein to ZmMYB-IF35 in wheat capable of regulating anthocyanin pathway [5].

## MATERIALS AND METHODS

### Target sequence

R2R3 MYB transcription factor MYB-IF35 (*Zea mays*) (NP\_001105092.1) protein sequence in FASTA format was obtained from NCBI (<https://ncbi.nlm.nih.gov/>).

### BLAST analysis

The sequence was subjected to BLASTp under *Triticum* (taxid: 4564) taxonomy. Among maximum E-value and score proteins, 'MYB-related protein P-like' (*Triticum dicoccoides*) (XP\_037430643.1) was selected with max score (247), query cover (46%), E-value (8e-79) and percent identity (69.09%).

### Putative 3D model prediction

The FASTA sequence of 'MYB-related protein P-like' (*Triticum dicoccoides*) (XP\_037430643.1) was search in Protein Data Bank (PDB) by uploading the NCBI ID in PDB. No results or sequence was found in PDB for the 'MYB-related protein P-like' (*Triticum dicoccoides*). So, the putative model was built through SWISS MODEL (<https://swissmodel.expasy.org/interactive>) by uploading the protein sequence of 'MYB-related protein P-like' in FASTA format from NCBI.

### Validation and quality assessment of putative protein models

The putative protein models were validated by 'VERIFY 3D' and quality was assessed through 'PROCHECK' by downloading protein model files from SWISS MODEL and uploading it to ERRAT tool. Visualization of the valid and best quality model was performed using PyMOL.

### Identification of physical and chemical properties

The amino acid sequence of 'MYB-related protein P-like' of the approved model (through SWISS MODEL) was uploaded in ExPASy\_ProtParam (<https://web.expasy.org/protparam/>) tool for the identification of physical and chemical parameters such as; molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity (GRAVY) [6].

### Identification of sub cellular location of 'MYB-related protein P-like'

For identifying sub cellular location of our protein, we uploaded the amino acid sequence of our putative protein in Protcom 9.0 tool.

### Secondary structure prediction

For determination of secondary structure of the MYB-related protein P-like anthocyanin transcription factor; its protein sequence in FASTA format was uploaded to PSIPRED (<http://bioinf.cs.ucl.ac.uk/psipred/>) and SOPMA (<https://npsa>

prabi.ibcp.fr/cgi-bin/npsa\_automat.pl? page=/NPSA/npsa\_sopma.html). Both tools accurately predict secondary structures accurately.

### Transmembrane helix topology

In PSIPRED, transmembrane helix prediction is done by MEMSATSV (http://bioinf.cs.ucl.ac.uk/psipred/) which is highly accurate predictor of transmembrane helix topology along with identification of cytosolic and extra-cellular loops present in protein structure.

### GO prediction: Biological, cellular and molecular function

The physio-chemical and gene ontology properties of 'MYB-related protein P-like' anthocyanin transcription factor was done by uploading the amino acid sequence of our putative protein in PSIPRED FFPred 3 tool (http://bioinf.cs.ucl.ac.uk/psipred/). The biological, cellular and molecular functions of our protein were analyzed.

**Post-translational modifications:** Phosphorylation, glycosylation and PEST regions present within the protein sequence were predicted by FFPred PSIPRED by uploading the amino acid sequence to PSIPRED web tool.

**Protein functional residues prediction:** Functional residues and available hotspots or residues in the protein were predicted through HotSpot Wizard 3 available at: https://loschmidt.chemi.muni.cz/hotspotwizard/. In this program, predictions are performed by combining the functional, evolutionary and structural data from various computational tools.

**Ligand binding residues prediction:** Active sites present within our protein of interest were determined by uploading sequence in FASTA format to IntFOLD server (version 7.0) (https://www.reading.ac.uk/bioinf/IntFOLD/).

**Genomic assessment:** Using NCBI, the gene for 'MYB-related protein P-like' protein was obtained through NCBI database. The gene was analyzed for locus of the gene in wheat, number of exons, gene symbol and conserved domains present in the gene sequence [7].

## RESULTS AND DISCUSSION

### Validation quality assessment of putative protein models

The putative model built through SWISS MODEL. Resulted in 4 hypothetical models. The PDB files of all the models were downloaded. In order to determine the valid 3D model for 'MYB-related protein P-like' (*Triticum dicoccoides*), ERRAT software was used. The downloaded PDB file of model-01 constructed through SWISS MODEL was uploaded and program was made to run. The model was subjected to 'VERIFY 3D' (to validate the 3D structure of the hypothetical model) and 'PROCHECK' (to check the quality of the model and residues present in the model) (Figure 3). The 3D model of model 01 was

"fail" since, 68.53% of the residues have averaged 3D-1D score  $\geq 0.2$  and fewer than 80% of the amino acids have scored  $\geq 0.2$  in the 3D/1D profile. However, structure contained one residue Ser 90(C) in disallowed region and one residue ASP 134 (C) in generously allowed region (Figure 7).

Similarly, the downloaded PDB file of model 02 constructed through SWISS MODEL, was uploaded and program was made to run. The model was subjected to 'VERIFY 3D' and 'PROCHECK'. The 3D structure of the model was "rejected" since, 38.89% of the residues have averaged 3D-1D score  $\geq 0.2$  and fewer than 80% of the amino acids have scored  $\geq 0.2$  in the 3D/1D profile (Figure 4). The quality of the structure determined through PROCHECK described that the structure had 4 errors and 1 warning. At least 90% of the residues should be in most favored region and no important residue in disallowed region. Some of the glycine residues were found in disallowed region (Figure 8).

Next, the downloaded PDB file of model 03 constructed through SWISS MODEL was uploaded and program was made to run. The model was subjected to 'VERIFY 3D' and 'PROCHECK'. 94% of the residues have averaged 3D-1D score  $\geq 0.2$ . Hence, the 3D structure was predicted to be "pass". At least 80% of the amino acids have scored  $\geq 0.2$  in the 3D/1D profile (Figure 5). The Ramachandran plot, also shows no amino acid residue in disallowed region (Figure 9). However, one of the amino acid is present in generously allowed region [8].

Lastly, the downloaded PDB file of model 04 constructed through SWISS MODEL, was uploaded and program was made to run. The model was subjected to 'VERIFY 3D' and 'PROCHECK'. 77.36% of the residues have the average 3D-1D score  $\geq 0.2$ , so the structure model was predicted to be "not valid". Moreover, fewer than 80% of the amino acids have the score  $\geq 0.2$  in the 3D/1D profile (Figure 6). The Ramachandran plot, also shows no amino acid residue in disallowed region but 7 of the amino acid are present in generously allowed region (Figure 10) [9].

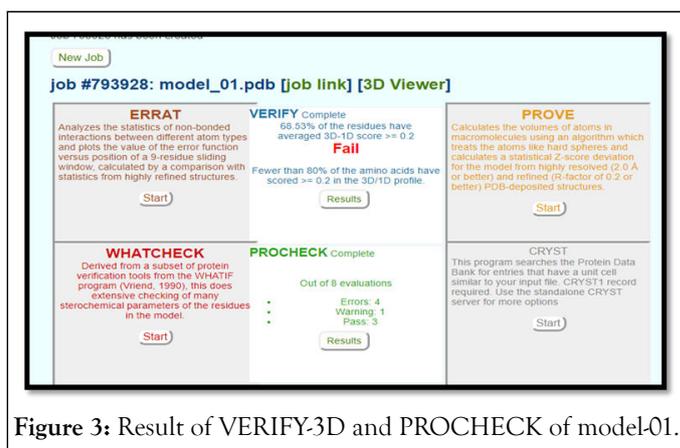


Figure 3: Result of VERIFY-3D and PROCHECK of model-01.

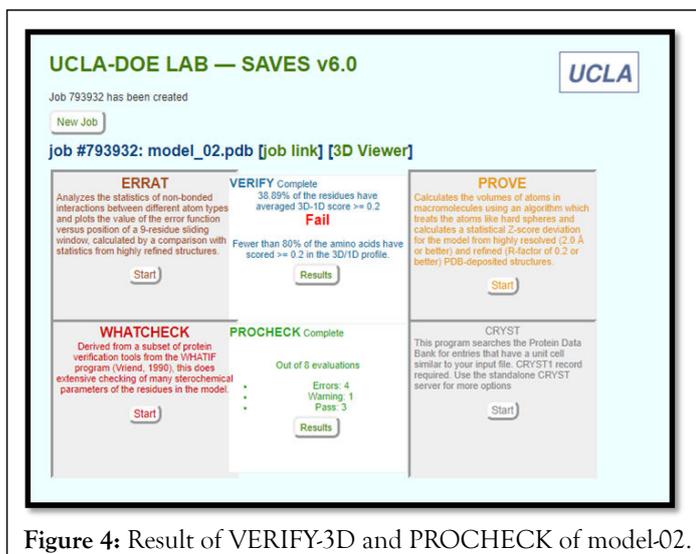


Figure 4: Result of VERIFY-3D and PROCHECK of model-02.

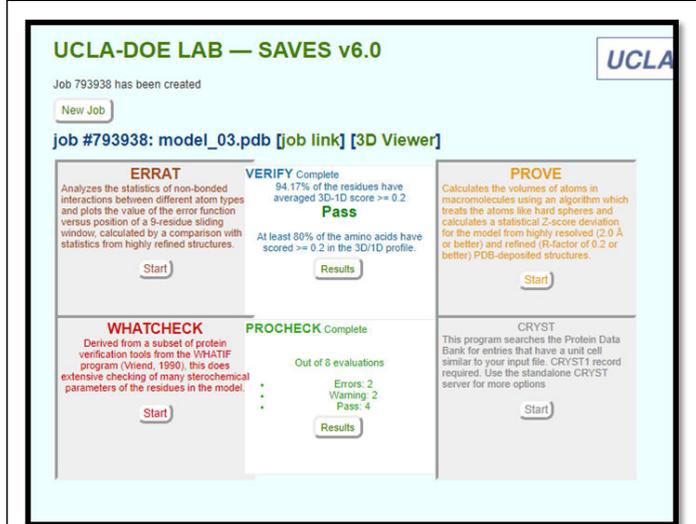


Figure 5: Result of VERIFY 3D and PROCHECK of model-03.

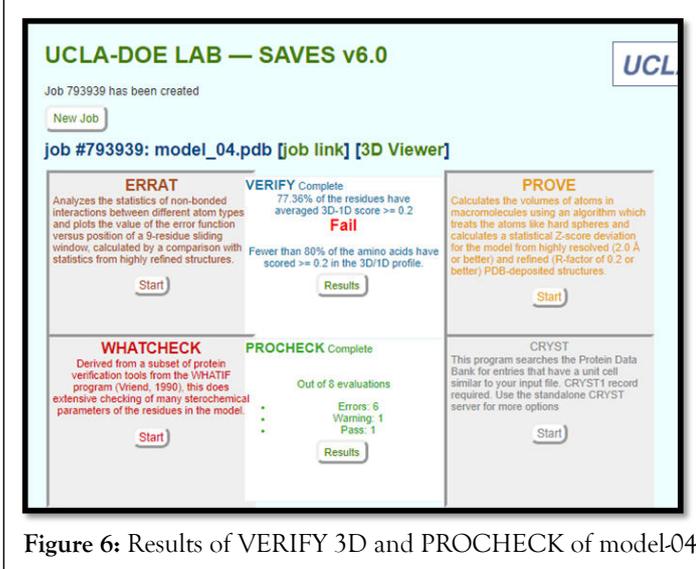


Figure 6: Results of VERIFY 3D and PROCHECK of model-04.

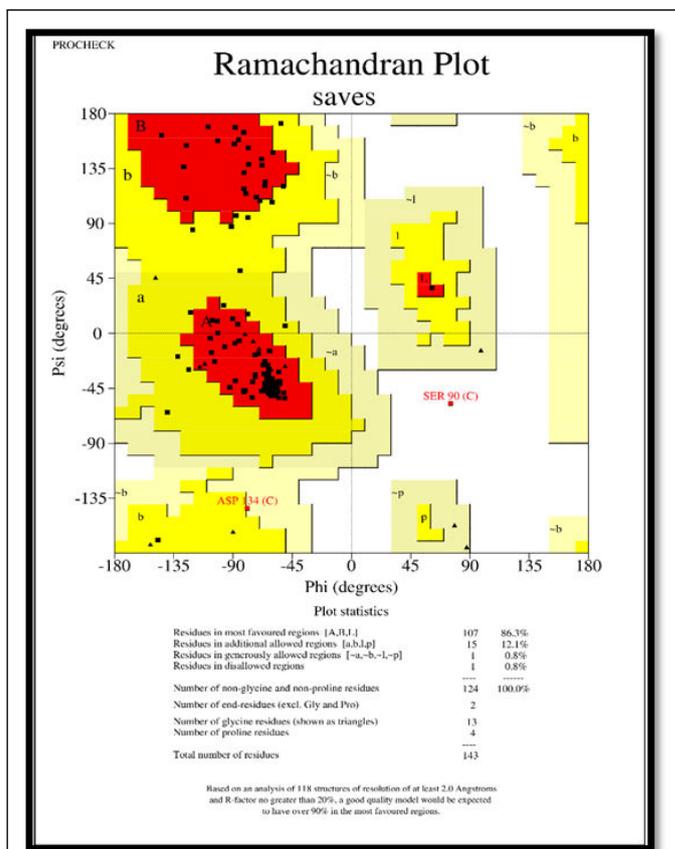


Figure 7: Ramachandran plot of model-01.

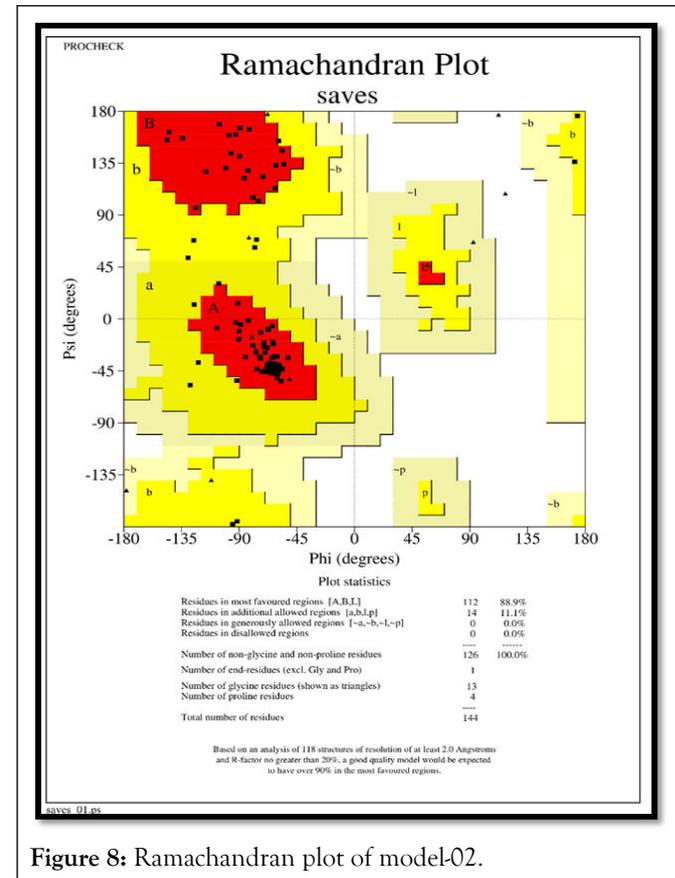


Figure 8: Ramachandran plot of model-02.

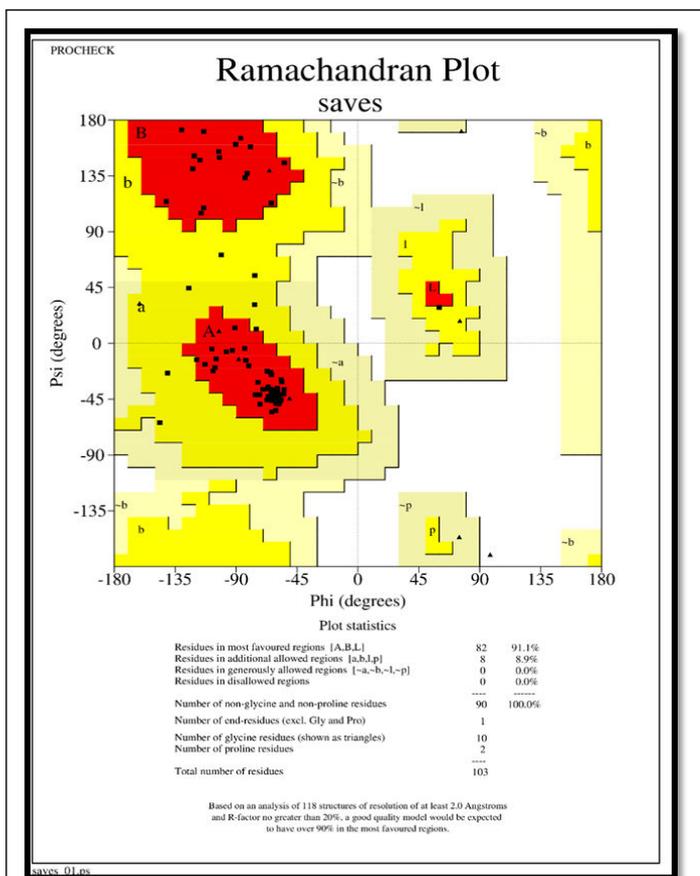


Figure 9: Ramachandran plot of model-03.

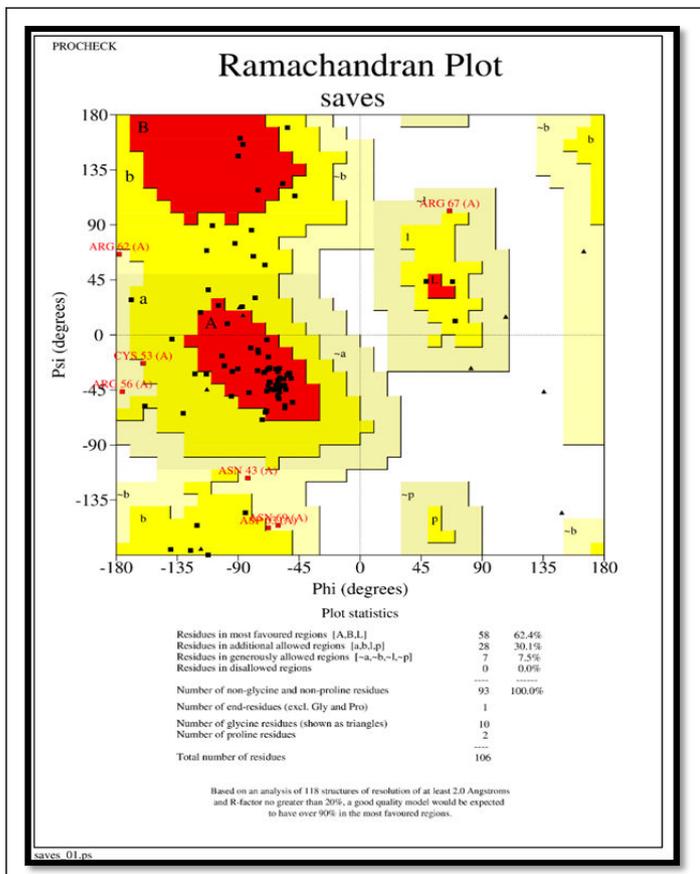


Figure 10: Ramachandran plot of model-04.

### Putative structure of 'MYB-related protein' transcription factor (*Triticum dicoccoides*)

The hypothetical structure of the 'MYB-related protein P-like' sequence (model-03) obtained by using SWISS-MODEL and validated by ERRAT, was then analyzed by using PyMOL for structural features. From the description of SWISS MODEL, it is evident that the model has 45.19% identity with the template sequence i.e., R2R3-MYB (*Zea mays*). The structure prediction was performed through X-ray crystallography and is of 1.68 Angstrom (Figure 11). The PDB file of model 3 was downloaded for visualization in PyMOL.

Predicted structure of 'MYB-related protein P-like' (*Triticum dicoccoides*) constructed by SWISS MODEL shows that the chain is continuous since no break was noticed and the molecule consists of only one chain i.e., CHAIN-A as shown in (Figures 12-14).

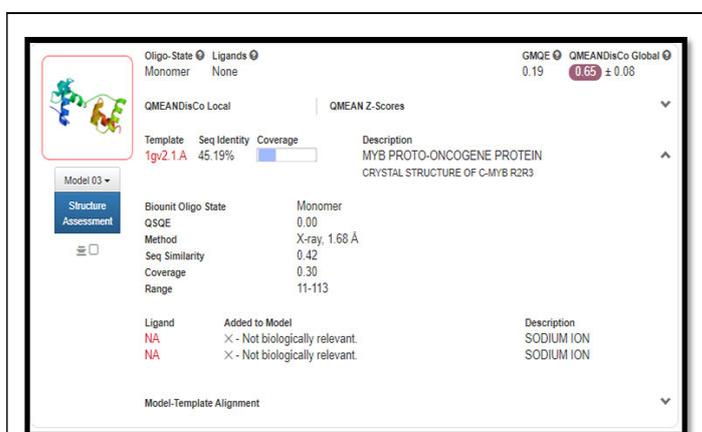


Figure 11: Description of predicted yet valid model-03 by SWISS MODEL.

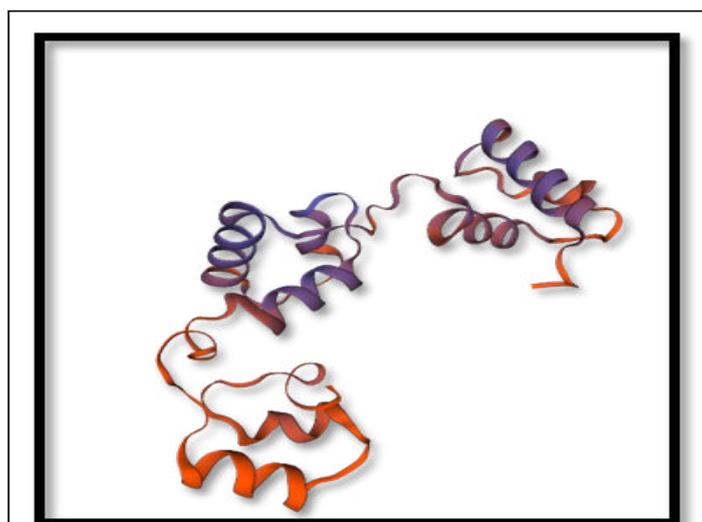


Figure 12: Predicted structure of MYB-related protein P-like' (*Triticum dicoccoides*) constructed by SWISS MODEL.

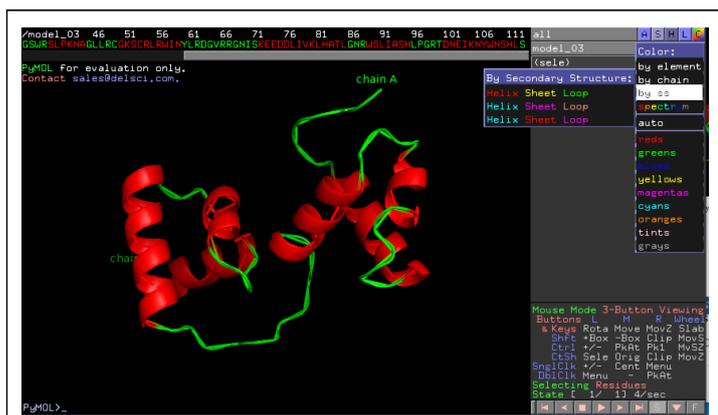


Figure 13: Sheaths and helices shown by PyMoL in MYB-related protein P-like' (*Triticum dicoccoides*).

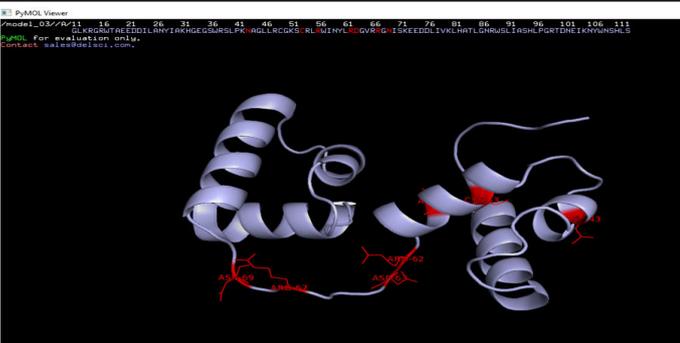


Figure 14: PyMoL showing amino acid residues residing in generously allowed region of MYB-related protein P-like' (*Triticum dicoccoides*) as determined by Ramachandran plot report through ERAAT, MYB-related protein P-like' (*Triticum dicoccoides*).

6 beta helix (shown by red), no sheath and 6 loops joining the helix structures (shown by green color) were found in the structure. ASN-69, ARG-67 and ARG-62 are the residues that are present in generously allowed region but far away from allowed region (Figure 14) but ARG-56, ASN-43 and CYS-53 are although shown by Ramachandran plot to be in generously allowed region but they are located very near to allowed region [10].

### Physical and chemical properties of 'MYB-related protein P-like' anthocyanin transcription factor

'MYB-related protein P-like' (C<sub>1618</sub>H<sub>2565</sub>N<sub>479</sub>O<sub>534</sub>S<sub>15</sub>) is composed of 346 amino acids with molecular weight of 37752.95 and isoelectric point (PI) of 5.19. the protein is overall negatively charged with 47 negatively charged residues (Asp+Glu) and relatively less positively charged amino acids (Arg+Lys) i.e., 35. Proteins with negative GRAVY index are hydrophilic so our protein was predicted to be hydrophilic with -0.595 GRAVY index. The protein was found to be thermally stable since its aliphatic index fall in the range of 66.5-84.33 i.e., 72.75 (Table 1 and Figure 15) [11].

Table 1: Expaty protparam 'MYB-related protein P-like' (*Triticum dicoccoides*) (XP\_037430643.1).

Number of amino acid	346
Molecular weight	37752.95
Theoretical pl	5.19
Amino acid composition	Ala (A) 25 7.2%
	Arg (R) 20 5.8%
	Asn (N) 17 4.9%
	Asp (D) 25 7.2%
	Cys (C) 7 2.0%
	Gln (Q) 15 4.3%
	Glu (E) 22 6.4%
	Gly (G) 28 8.1%
	His (H) 8 2.3%
	Ile (I) 16 4.6%
	Leu (L) 28 8.1%

	Lys (K) 15 4.3%
	Met (M) 8 2.3%
	Phe (F) 3 0.9%
	Pro (P) 19 5.5%
	Ser (S) 40 11.6%
	Thr (T) 17 4.9%
	Trp (W) 10 2.9%
	Tyr (Y) 4 1.2%
	Val (V) 19 5.5%
	Pyl (O) 0 0.0%
	Sec (U) 0 0.0%
	(B) 0 0.0%
	(Z) 0 0.0%
	(X) 0 0.0%
Total number of negatively charged residues (Asp+Glu)	47
Total number of positively charged residues (Arg+Lys)	35
Atomic composition	Carbon C 1618
	Hydrogen H 2565
	Nitrogen N 479
	Oxygen O 534
	Sulfur S 15
Formula	C <sub>1618</sub> H <sub>2565</sub> N <sub>479</sub> O <sub>534</sub> S <sub>15</sub>
Total number of atoms	5211
Extinction coefficients	Extinction coefficients are in units of M <sup>-1</sup> cm <sup>-1</sup> , at 280 nm measured in water.
	Ext. coefficient 61335 Abs 0.1% (=1 g/l) 1.625, assuming all pairs of Cys residues form cystines
	Ext. coefficient 60960 Abs 0.1% (=1 g/l) 1.615, assuming all Cys residues are reduced
Estimated half-life	The N-terminal of the sequence considered is M (Met)
	The estimated half-life is: 30 hours (mammalian reticulocytes, <i>in vitro</i> )
	>20 hours (yeast, <i>in vivo</i> )

	>10 hours ( <i>Escherichia coli</i> , <i>in vivo</i> )
Instability index	The instability index (II) is computed to be 54.80
	This classifies the protein as unstable
Aliphatic index	72.75
Grand average of hydropathicity (GRAVY)	-0.595

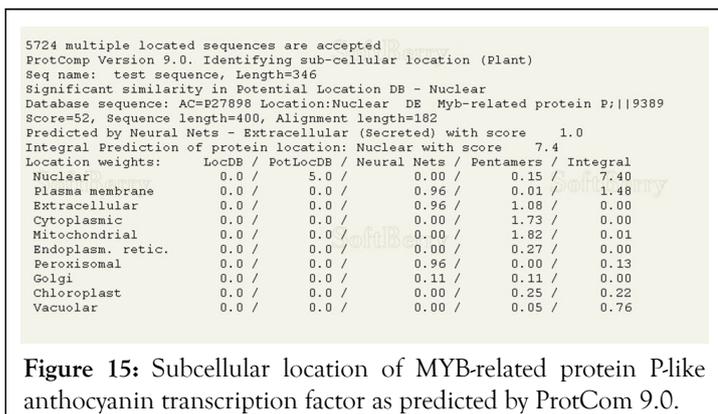


Figure 15: Subcellular location of MYB-related protein P-like anthocyanin transcription factor as predicted by ProtCom 9.0.

(Here; 'LocDBare' scores based on query protein's homologies with proteins of known localization. 'PotLocDBare' scores based on homologies with proteins which locations are not experimentally known but are assumed from strong theoretical evidence. 'Neural Nets' are scores assigned by neural networks. 'Pentamers' are scores based on comparisons of pentamer distributions calculated for QUERY and DB sequences. 'Integral' is final scores that combine all above previous scores. The scores are renormalized to have a sum of scores for all localizations equal 3 (Nnets), 5 (PotLocDB, Pentamers) or 10 (LocDB, Integral)).

From protcom 9.0, the subcellular location of myb-related protein P-like anthocyanin transcription factor was found to be nuclear with ProtLocDB score (scores based on homologies with proteins which locations are not experimentally known but are assumed from strong theoretical evidence) of 5.0 (Figure 15). According to the neural net scores, the protein was predicted to be "secreted to extracellular space" with scores of almost 1.0 (i.e., 0.96) [12].

### Secondary structure prediction

Protein secondary structures predicted by SOPMA plots predicted that random coils consisted of the most significant portion of the protein (50%), alpha helix placed at second (36.99%), extended rows placed at third (8.38%) and beta turns placed at last (4.62%) (Figure 16 and Table 2). On the other hand, secondary structure prediction by PSIPRED, showed alpha helix (34%) and coils (66%) (Figure 17). PSIPRED also demonstrated the existence of small nonpolar (39%), hydrophobic (19%),

Table 2: Calculated secondary structures by SOPMA plots.

### MYB-predicted protein P-like

polar (34%) and aromatic plus cysteine residues (6%) (Figure 18) [13].

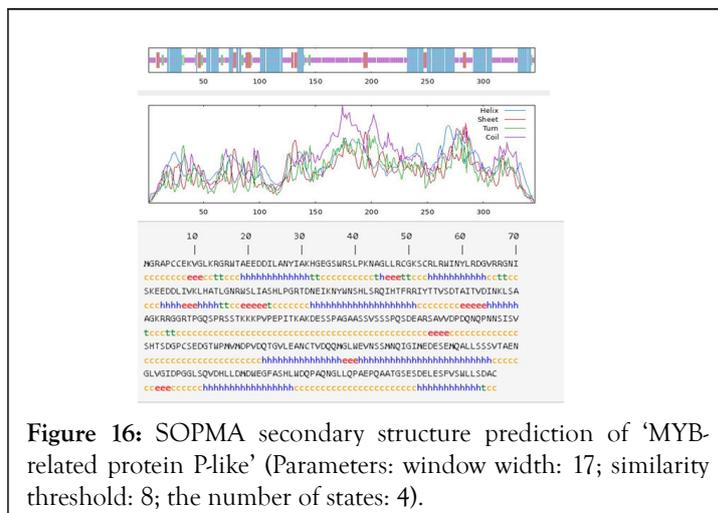


Figure 16: SOPMA secondary structure prediction of 'MYB-related protein P-like' (Parameters: window width: 17; similarity threshold: 8; the number of states: 4).

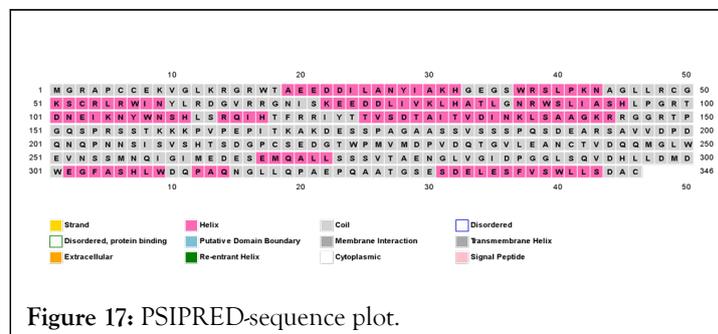


Figure 17: PSIPRED-sequence plot.

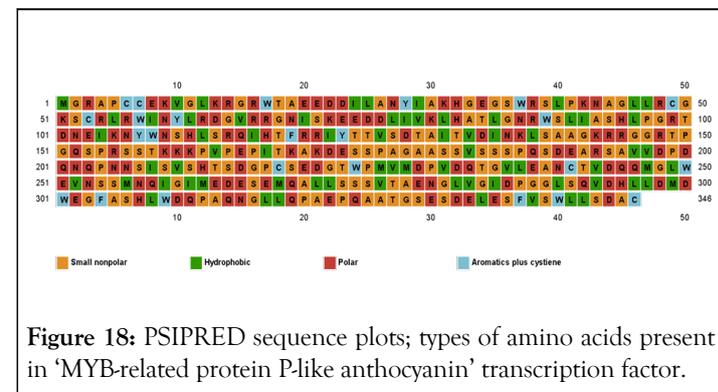
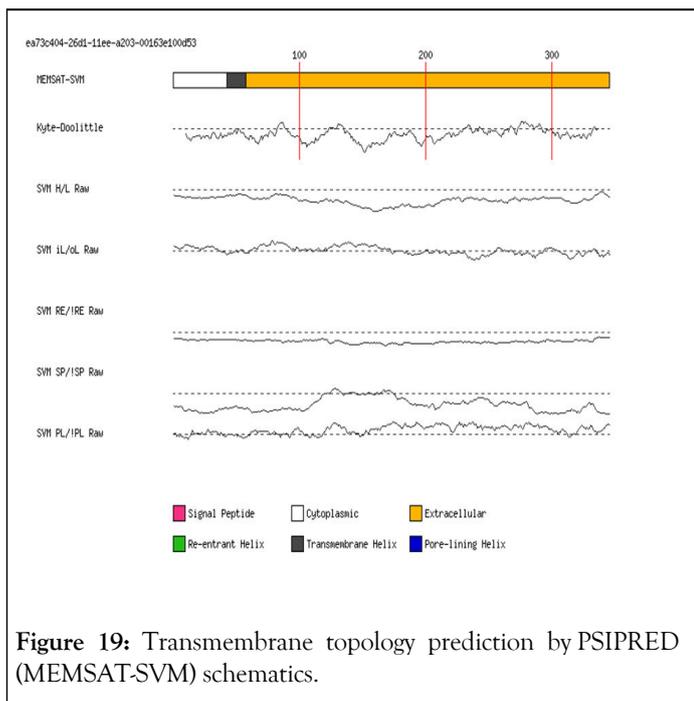


Figure 18: PSIPRED sequence plots; types of amino acids present in 'MYB-related protein P-like anthocyanin' transcription factor.

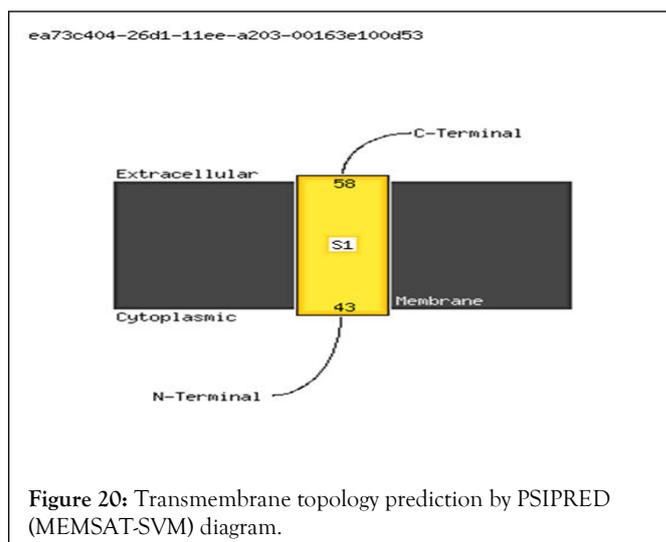
Number of amino acids	346
$\alpha$ -helix	36.99%
$\beta$ -turn	4.62%
Extended rows	8.38%
Random coils	50.00%

### Transmembrane topology

Transmembrane topology prediction by PSIPRED (MEMSAT-SVM) showed that out of 346 amino acids present in the transcription factor, few amino acids at beginning (Met1-Lys42) of the sequence are immersed in cytoplasmic region with very few next amino acids contributing towards a transmembrane helix (Asn43-Ile58). Rest of the portion of the sequence belongs is majorly extracellular (Asn59-Cys346). No signal peptide, re-entrant helix and pore-lining helix were found in the structure. Figure 19, shows the predicted topology of the protein structure with its diagrammatic representation (Figure 20) [14].



**Figure 19:** Transmembrane topology prediction by PSIPRED (MEMSAT-SVM) schematics.



**Figure 20:** Transmembrane topology prediction by PSIPRED (MEMSAT-SVM) diagram.

### GO prediction: Biological, cellular and molecular function

Gene ontology and predictors of physio chemical properties of 'MYB-related protein P-like anthocyanin transcription factor through FFPred 3 tool predicted all GO domains *i.e.*, biological process, molecular function predictions and cellular component predictions. Biological function of our protein was predicted to be bind regulation of RNA biosynthetic process (GO:2001141) with highest probability (0.944), regulation of gene expression (GO:0010468) with second highest probability (0.937) and cellular macromolecule biosynthetic process (GO: 0051171) with third highest probability (0.934) (Table 3). The prediction of molecular function of our putative protein confirms that it is a transcription factor for the regulation RNA biosynthesis of anthocyanin pathway [15].

**Table 3:** Predicted biological functions through PSIPRED (FFPred).

Biological processes			
GO term	Name	Probability	SVM (Support Vector Machine) reliability
GO:2001141	Regulation of RNA biosynthetic process	0.944	H
GO:0010468	Regulation of gene expression	0.937	H
GO:0034645	Cellular macromolecule biosynthetic process	0.934	H
GO:0051171	Regulation of nitrogen compound metabolic process	0.931	H
GO:0019222	Regulation of metabolic process	0.92	H
GO:1903506	Regulation of nucleic acid-templated transcription	0.914	H
GO:0006355	Regulation of transcription, DNA-templated	0.881	H
GO:0051252	Regulation of RNA metabolic process	0.876	H
GO:0006351	Transcription, DNA-templated	0.853	H
GO:0009059	Macromolecule biosynthetic process	0.844	H
GO:0006357	Regulation of transcription from RNA polymerase II promoter	0.772	H
GO:0006810	Transport	0.737	H
GO:0010629	Negative regulation of gene expression	0.737	H
GO:0006397	mRNA processing	0.73	H
GO:0000375	RNA splicing, <i>via</i> transesterification reactions	0.696	H
O:0009890	Negative regulation of biosynthetic process	0.67	H
GO:0045892	Negative regulation of transcription, DNA-templated	0.669	H
GO:0008380	RNA splicing	0.655	H
GO:0006366	Transcription from RNA polymerase II promoter	0.628	H
GO:0006396	RNA processing	0.611	H
GO:0051641	Cellular localization	0.585	H

GO:0031328	Positive regulation of cellular biosynthetic process	0.549	H
GO:0010628	Positive regulation of gene expression	0.537	H
GO:0045944	Positive regulation of transcription from RNA polymerase II promoter	0.533	H
GO:0051649	Establishment of localization in cell	0.522	H

The protein was predicted to be a component of nucleolus (GO:0005730) with highest probability (0.696), nuclear body (GO:0016604) with second highest probability (0.662), ribonucleoprotein complex (GO:0030529) with third highest probability (0.625) and plasma membrane with least high probability (0.514) (Table 4).

**Table 4:** Predicted cellular component through PSIPRED (FFPred).

Cellular component predictions			
GO term	Name	Probability	SVM (Support Vector Machine) reliability
GO:0005730	Nucleolus	0.696	H
GO:0016604	Nuclear body	0.662	H
GO:0030529	Ribonucleoprotein complex	0.625	H
GO:0005886	Plasma membrane	0.514	H

Molecular function of our protein was predicted to be bind to nucleic acids (GO:0003676), DNA (GO:0003677), zinc ions (GO:0008270), sequence-specific DNA binding transcription factor activity (GO:0003700), transition metal ion binding (GO:0046914), cytoskeletal protein binding (GO:0008092), nucleic acid binding transcription factor activity (GO:0001071), kinase binding (GO:0019900), catalytic activity (GO:0003824), protein heterodimerization activity (GO:0046982), actin binding (GO:0003779), RNA binding (GO:0003723), microtubule binding (GO:0008017), sequence-specific DNA binding RNA

polymerase II transcription factor activity (GO:0000981) and protein kinase binding (GO:0019901) with probabilities ranging from highest to lowest respectively (Table 5).

However, the highest molecular function probability of 'MYB-related protein P-like' was predicted to be nucleic acid binding (probability; 0.974).

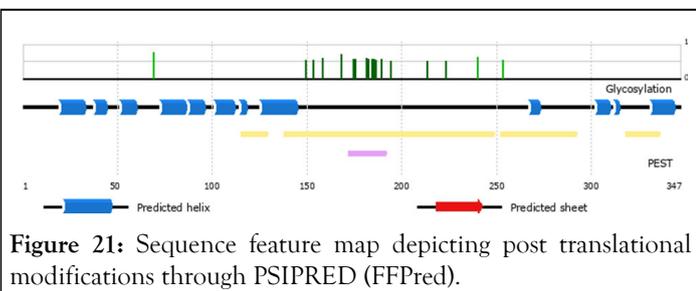
**Table 5:** Predicted molecular function through PSIPRED (FFPred).

Molecular function prediction			
GO term	Name	Probability	SVM (Support Vector Machine) reliability
O:0003676	Nucleic acid binding	0.974	H
GO:0003677	DNA binding	0.941	H
GO:0008270	Zinc ion binding	0.91	H
GO:0003700	Sequence-specific DNA binding transcription factor activity	0.881	H
GO:0046914	Transition metal ion binding	0.827	H

GO:0008092	Cytoskeletal protein binding	0.802	H
GO:0001071	Nucleic acid binding transcription factor activity	0.787	H
GO:0019900	Kinase binding	0.772	H
GO:0003824	Catalytic activity	0.738	H
GO:0046982	Protein heterodimerization activity	0.708	H
GO:0003779	Actin binding	0.661	H
GO:0003723	RNA binding	0.594	H
GO:0008017	Microtubule binding	0.589	H
GO:0000981	Sequence-specific DNA binding RNA polymerase II transcription factor activity	0.553	H
GO:0019901	Protein kinase binding	0.535	H

### Prediction of post-translational modifications

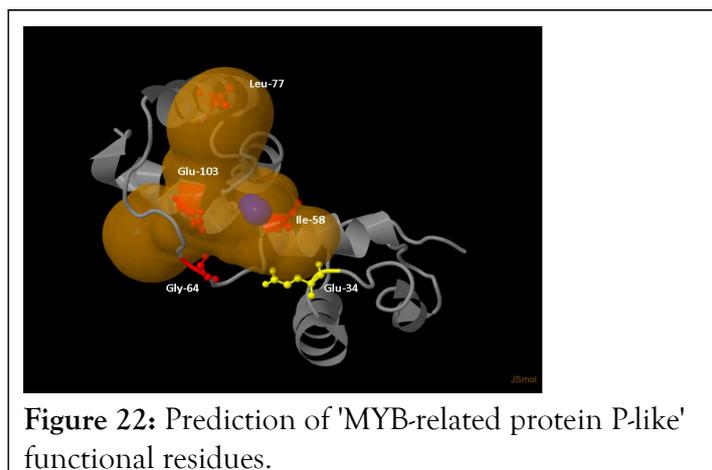
Phosphorylation, glycosylation and PEST regions present within the protein sequence were predicted by FFPred PSIPRED. The sequence feature map below shows the glycosylation with green lines and the line height indicates the confidence of the residue prediction. No phosphorylation was predicted onto the sequence. PEST domain in protein sequence represents sequence rich in Proline (P), Glutamic acid (E), Serine (S) and Threonine (T). These sequences are flanked by positively charges residues such as lysine (K), arginine (R) and histidine (H). PEST sequences play a role in displaying a proteolytic signal towards target protein for its degradation resulting in short intracellular half-lives. The sequence feature map of FFPred predicts presence of PEST region in our protein residing in between Pro150-Asp200 as shown in purple. Moreover, the most of the amino acids were also predicted to be glycosylated between pro150 to Asp200 (Figure 21) [16].



### Protein functional residues prediction

The files of 'MYB related protein P-like' were imported into the HotSpot Wizard 3 server. The amino acids Ile-58, Gly-64, Leu-77, Glu-103 and Glu-34 were predicted as highly reliable and mutable residues situated at the catalytic pockets. Gly-64,

Leu-77, Glu-103, Glu-34 were predicted outside the tunnel (Figure 22).



**Figure 22:** Prediction of 'MYB-related protein P-like' functional residues.

### Predicted ligand binding residues

Active sites present within our protein of interest were determined by uploading sequence in FASTA format to IntFOLD server (version 7.0). In MYB-related protein P-like *Triticum dicoccoides*, amino acids at 180-198 and 225-228 were predicted to be involved in binding dopamine (DA382; DA893) and ruthenium (RU914) ligands most likely. The nature of attraction was predicted to be DNA binding (selectively and non-covalently) (GO term; 0003677) and binding to chromatin, the network of fibers of DNA, protein and sometimes RNA, that make up the chromosomes of the eukaryotic nucleus during interphase (GO term; 0003682) (Table 6).

**Table 6:** Predicted ligand binding residues through IntFOLD server (version 7.0).

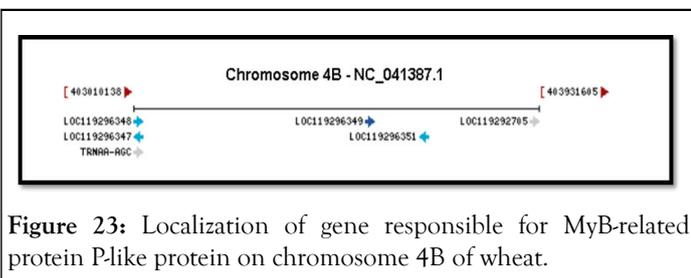
Binding site	180, 181, 182, 183, 184; 225, 226, 227, 228	
Most likely ligands at each site (Type)	DA; RU; DA	Dopamine (DA), Ruthenium (RU)
Centroid ligands at each site (TypeID)	DA382; RU914; DA893	
All ligands in clusters (Type-frequency)	RU-7, DC-41, DA-53, DT-51, DG-38; RU-4; DC-2, DA-3, DG-1	
Likely+centroid ligands at each site	DA382; RU914; DA893	
GO terms	003677	DNA binding molecular function any molecular function by which a gene product interacts selectively and non-covalently with DNA (Deoxyribonucleic Acid).
GO terms	003682	Chromatin binding gene ontology term (GO: 0003682) Definition: Binding to chromatin, the network of fibers of DNA, protein and sometimes RNA, that make up the chromosomes of the eukaryotic nucleus during interphase. Molecular function

**Genomic assessment**

Assessment of gene responsible for the transcription of protein sequence of our protein ‘Myb-related protein P-like’ was performed through NCBI database. The gene is located on

**Table 7:** Genomic characterization through NCBI.

Sr. no	Genomic characterization of MyB-related protein P-like	
1	Locus (Gene symbol)	LOC119296349
2	Gene location	4B
3	Exons	3
4	Location	NC_041387.1 (403535222..403540492)



**Figure 23:** Localization of gene responsible for MyB-related protein P-like protein on chromosome 4B of wheat.

**Conserved domains**

Conserved domains present within a gene allows detection of polypeptide sequences due to presence of conserved patterns and motifs. Through NCBI-conserved domain database, we found specific hits, non-specific hits and super-families of domains present within the sequence. Among specific hits, Arg-99, Thr-100, Asn-102, Glu-103, Lys-105, Asn-106, Tyr-107, Asn-109, Ser-110 and His-112 were found to be involved in playing role in DNA binding capability of ‘MYB-related protein P-like’ protein.

chromosome 4B on LOC119296349 locus and contains 3 exons (Table 7 and Figure 23).

Moreover, these nucleotides involved in transcribing these amino acids in this domain were also reported to be involved in chromatin remodeling (Transcription, DNA repair, replication and recombination) through interaction with histones. However, a total of 14 conserved domains were found in the sequence.

A total of three superfamilies PLN03212 (transcription repressor) REB1 superfamily (Myb superfamily proteins, including transcription factors and mRNA splicing factors) and SANT superfamily (DNA binding as well as capping complex) were predicted near N-terminal of sequence. Only SANT superfamily got ‘specific hits’ while other two superfamilies got ‘non-specific hits’.

In PLN0321; a transcription repressor MYB5 superfamily only one PLN03212 domain (Transcription repressor MYB5; Provisional) was predicted that has the capability to bind to promoters of the DNA sequence and delay binding of RNA polymerases.

In REB1 superfamily, REB1, MYB\_DNA-binding, MYB\_DNA-bind\_6, SANT, SANT, My\_DNA-bind\_6, REB1 and SANT\_TRF domains were predicted.

In SANT superfamily, MYB\_DNA-binding, SANT, SANT domains were predicted. SANT superfamily domain was also found near N-terminal of the sequence which was predicted to be involved in overall DNA binding as well as capping complex (Figure 24 and Table 8). These results will help in the purification of our protein in future for its physiochemical and biological functions in wheat.

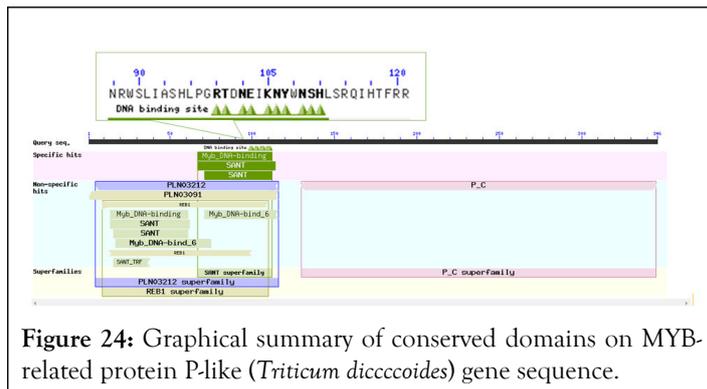


Figure 24: Graphical summary of conserved domains on MYB-related protein P-like (*Triticum dicccoides*) gene sequence.

Table 8: List of conserved domain hits on MYB-related protein P-like (*Triticum dicccoides*) gene sequence.

Name	Accesion	Description	Interval	Evalue
PLN03212 super family	cl31985	Transcription repressor MYB5; Provisional	5-116	6.27e-57
P_C super family	cl05925	P protein C-terminus; This family represents the C-terminus of plant P proteins. The maize P gene is a transcriptional regulator of genes encoding enzymes for flavonoid biosynthesis in the pathway leading to the production of a red phlobaphene pigment and P proteins are homologous to the DNA-binding domain of MYB-like transcription factors. All members of this family contain the pfam00249 domain	130-345	4.19e-44
PLN03091	PLN03091	Hypothetical protein; Provisional	1-116	5.80e-54
MYB_DNA-binding	pfam00249	Myb-like DNA-binding domain; This family contains the DNA binding domains from Myb proteins, as well as the SANT domain family	67-112	2.67e-17
REB1	COG5147	Myb superfamily proteins, including transcription factors and mRNA splicing factors.	9-110	9.14e-14
SANT	smart00717	SANT SW13, ADA2, N-CoR and TFIIIB" DNA-binding domains	67-114	1.19e-13
Myb_DNA-binding	pfam00249	Myb-like DNA-binding domain; This family contains the DNA binding domains from Myb	14-61	5.69e-13

		proteins, as well as the SANT domain family		
SANT	cd00167	'SWI3, ADA2, N-CoR and TFIIB' DNA-binding domains. Tandem copies of the domain bind telomeric DNA tandem repeats as part of the capping complex	71-112	8.99e-13
SANT	smart00717	SANT SWI3, ADA2, N-CoR and TFIIB" DNA-binding domains	14-62	7.74e-11
SANT	cd00167	'SWI3, ADA2, N-CoR and TFIIB' DNA-binding domains. Tandem copies of the domain bind telomeric DNA tandem repeats as part of the capping complex.	16-61	1.28e-10
Myb_DNA-bind_6	pfam13921	Myb-like DNA-binding domain; This family contains the DNA binding domains from Myb proteins, as well as the SANT domain family	17-75	2.71e-09
Myb_DNA-bind_6	pfam13921	Myb-like DNA-binding domain; This family contains the DNA binding domains from Myb proteins, as well as the SANT domain family	71-116	3.90e-09
REB1	COG5147	Myb superfamily proteins, including transcription factors and mRNA splicing factors	13-100	2.10e-03
SANT_TRF	cd11660	Telomere repeat binding factor-like DNA-binding domains of the SANT/myb-like family	16-38	8.34e-03

## Future prospects

By utilizing this study, suitable ligands can be determined through molecular docking that can be utilized in for genetic engineering of the bio-synthetic pathway of anthocyanins in wheat. The study is utilized for the isolation of predicted protein and its identification for its homologues in wheat and its role in various biochemical processes particularly anthocyanin biosynthetic pathway. Moreover, this study can also be used for enhancement of anthocyanin pathway genetically in wheat and identification of its capability to combat various environmental stresses.

## CONCLUSION

The present study shows that hypothetical prediction of anthocyanin transcription factor is essential for its isolation, synthesis and characterization. The characterization of hypothetical transcription factor 'MYB-related protein P-like' in wheat (*Triticum diccoides*) with maximum similarity to R2R3-Myb transcription factor in *Zea mays* was completed by employing various bioinformatics tools which led to the novel characterization of transcription factor. This method takes minimums time, is more cost effective and provides additional benefit for further purification and isolation of the protein of interest.

## REFERENCES

1. Abdel-Aal ES, Hucl P. Composition and stability of anthocyanins in blue-grained wheat. *J Agric Food Chem.* 2003;51(8):2174-2180.
2. Ambawat S, Sharma P, Yadav NR, Yadav RC. MYB transcription factor genes as regulators for plant responses: An overview. *Physiol Mol Biol Plants.* 2013;19:307-321.
3. Chen C, Zhang K, Khurshid M, Li J, He M, Georgiev MI, et al. MYB transcription repressors regulate plant secondary metabolism. *CRC Crit Rev Plant Sc.* 2019;38(3):159-170.
4. Costantini L, Malacarne G, Lorenzi S, Troglio M, Mattivi F, Moser C, et al. New candidate genes for the fine regulation of the colour of grapes. *J Exp Bot.* 2015;66(15):4427-4440.
5. Gould KS. Nature's Swiss army knife: The diverse protective roles of anthocyanins in leaves. *J Biotechnol.* 2004;2004(5):314.
6. Heine GF, Malik V, Dias AP, Grotewold E. Expression and molecular characterization of ZmMYB-1F35 and related R2R3-MYB transcription factors. *Mol Biotechnol.* 2007;37:155-164.
7. Hughes NM, Connors MK, Grace MH, Lila MA, Willans BN, Wommack AJ. The same anthocyanins served four different ways: Insights into anthocyanin structure-function relationships from the wintergreen orchid, *Tipularia discolor*. *Plant Sci.* 2021;303:110793.
8. Jiang W, Liu T, Nan W, Jeewani DC, Niu Y, Li C, et al. Two transcription factors TaPpm1 and TaPpb1 co-regulate anthocyanin biosynthesis in purple pericarps of wheat. *J Exp Bot.* 2018;69(10):2555-2567.
9. Lai KP, Dolan KD, Ng PK. Inverse method to estimate kinetic degradation parameters of grape anthocyanins in wheat flour under simultaneously changing temperature and moisture. *J Food Sci.* 2009;74(5):E241-E249.
10. Meng C, Sui N. Overexpression of maize MYB-1F35 increases chilling tolerance in arabidopsis. *Plant Physiol Biochem.* 2019;135:167-173.
11. Nielsen SL, Simonsen AM. Photosynthesis and photoinhibition in two differently coloured varieties of *Oxalis triangularis*-the effect of anthocyanin content. *Photosynthetica.* 2011;49:346-352.
12. Rojo F. Mechanisms of transcriptional repression. *Curr Opin Microbiol.* 2001;4(2):145-151.
13. Shin DH, Choi MG, Kang CS, Park CS, Choi SB, Park YI. A wheat R2R3-MYB protein purple plant1 (TaPL1) functions as a positive regulator of anthocyanin biosynthesis. *Biochem Biophys Res Commun.* 2016;469(3):686-691.
14. Tereshchenko O, Gordeeva E, Arbuzova V, Borner A, Khlestkina E. The D genome carries a gene determining purple grain colour in wheat. *Cereal Res Commun.* 2012;40(3):334-341.
15. Xie S, Lei Y, Chen H, Li J, Chen H, Zhang Z. R2R3-MYB transcription factors regulate anthocyanin biosynthesis in grapevine vegetative tissues. *Front Plant Sci.* 2020;11:527.
16. Yan H, Pei X, Zhang H, Li X, Zhang X, Zhao M, et al. MYB-mediated regulation of anthocyanin biosynthesis. *Int J Mol Sci.* 2021;22(6):3103.