I J T A
© Serials Publications
National Academy of Agricultural Science (NAAS) Rating : 3.03

Structural and Phylogenetic Analysis of Wheat AP2 Transcription Factor

A. K. Mishra, Neha Saini, Ohika Chakraborty, Chetna Tyagi, Amrender Kumar and A. K. Jain

ABSTRACT: *AP2 transcription factors play a critical role in plant development and adaptation to abiotic stress conditions. According to the number of AP2 domains, and the presence of other DNA binding domains, AP2/ERF can be divided into the ERF, AP2, RAV and Soloist families. This study aims to understand the evolutionary relationship, secondary structure and three-dimensional structure of AP2 protein through in silico approaches. We also study evolutionary relationships between AP2 transcription factors of wheat and other 10 plant species by generating a phylogenetic tree with the help of Maximum Likelihood method (ML). A consensus of AP2 sequences was used to model the protein. Based on structure prediction and analysis, it was inferred that the protein consists of a common super-secondary motif of sheet-helix-sheet type. The residues belonging to this region were found to have positive G-score value for phi-psi as well as all dihedral angles. A positive G-score value confers high probability conformation to the structure. Further investigation showed that the modeled three dimensional structure of AP2 was of high quality, as supported by the Ramachandran plot.*

Keywords*: AP2 domain, Maximum Likelihood, Phylogenetic analysis, Structure modeling*

INTRODUCTION

The common wheat or bread wheat (*T. aestivum*), an hexaploid species, is the most widely cultivated wheat species around the world. Wheat belongs to the genus *Triticum* and family *Poaceae*. The diploid (2n=7x=14) and tetraploid (2n=7x=42) species of the genus *Triticum* are cultivated. Abiotic stress includes any environmental conditions or combination of them that negatively affect the expression of genetic potential for growth, development and reproduction [1]. Wheat yields are depressed, among other factors, by drought, heat, low, temperatures, low fertility, especially nitrogen, and soil salinity. Economics, as well as, ecological limitation associated with these practices, however, have prompted an interest in searching for plant genetics resistance to environmental stress [2].

Environmental stresses can distort plant sustainability and productivity, which as a result involve activation of many stress responsive genes that help plants to combat and surpass the unfavourable environmental conditions [3]. These genes are divided into three major categories: (i) those that are involved in signaling pathways and in transcriptional control, such as MAP kinases and transcriptional factors (TFs) such as Heat shock factor (HSF) and dehydration responsive element-binding protein (DREB), basic leucine zipper (bZIP), MYB [4] etc. They interact with specific *cis*-regulatory element and increases expression of stress-regulated genes (ii) genes which are directly involved in the protection of membranes and proteins, such as heat shock proteins (HSPs), late embryogenesis abundant (LEA) proteins [2] and (iii) proteins that are involved in the uptake and transport of water and mineral ions across the cellular membrane for example aquaporin and ion transporters [5].

The APETALA 2 (*AP2)*, one of the largest groups of transcription factors, contains AP2 domains and present distinctively in plants. AP2 domain sequence consists of an approximate 60-80 amino acid residue that forms a characteristic alpha helix-turn-helix structure, responsible for DNA binding [6]. Based on number of AP2 domain, AP2 family is divided into three subclasses; (i) EREBPs and DREB contains single AP2 repeat with WLG motif (ii) RAV with two DNA binding motifs: a single AP2 repeat and a B3-like domain involved in regulating gene expression in response to ethylene, biotic and abiotic stresses (iii) DRE (dehydration-responsive element) are positive regulators of ABA signaling as determined from its effect on the ABA responsive gene [7, 8].

* AKMU, IARI, New Delhi-110012

In the present study, molecular modeling of AP2 protein has been carried out using the homology modeling approach to obtain the 3D structure. Eleven AP2 protein sequences were retrieved and their phylogenetic relationships have also been studied. The study also provides necessary and important information for understanding the molecular and phylogenetic basis of wheat AP2 protein form and function in other plant species. It also provides some insight into the possible function of this protein, thus, giving some directions for future functional analysis of AP2 transcription factor.

MATERIALS AND METHODS

Data Retrieval and Bioinformatics analysis

Transcription factor AP2, being our focus in different plant species, were retrieved through NCBI's Entrez [9] as a pre-requisite for phylogenetic analysis and structure modelling. The accession number of transcription factor AP2 in different species is given below (Table 1). These sequences were subjected to a multiple alignment analysis using ClustalW tool [10]. It is a progressive multiple sequence alignment program available either as a stand-alone or online program. The resultant multiple sequence alignement was used to infer a consensus AP2 sequence which was further used to model the three-dimensional structure.

Table 1
Protein Sequences and their and their Accession IDs

S.No.	Protein	Protein ID
1	*Mentha x piperita*	388565097
2	*Papaver somniferum*	388565095
3	*Brassica carinata*	81022813
4	*Brassica juncea*	81022811
5	*Arabidopsis arenosa*	38260685
6	*Triticum aestivum*	229002388
7	*Arabidopsis lyrata subsp*	297317115
8	*Olimarabidopsis pumila*	38260669
9	*Capsella rubella*	38260649
10	*Sisymbrium irio*	38260618
11	*Cocos nucifera*	3466549630

Phylogenetic analysis

The phylogenetic analysis by maximum likelihood was constructed with MEGA 6 software (Molecular Evolutionary Genetic Analysis) [11]. In order to construct the phylogeny, the pairwise alignment parameters were set with gap opening penalty 10; gap extension penalty 1.0; protein weight matrix BLOSUM 30. The number of bootstrap replications was set at 500 so that to obtain and compare an overall idea of the evolutionary relationship between the eleven plant species.

Modeling and validation of the protein structure

The 3D model for the AP2 proteins were built by I-Tasser [12]. I-TASSER server is an on-line platform for protein structure and function predictions. 3D models are built based on multiple-threading alignments. The conformational stability of the models was predicted by PSVS web server [13]. The models were visualized using Chimera [14].

RESULTS AND DISCUSSION

Multiple alignment of the AP2 protein in wheat and other plant species

The protein sequences were retrieved from the public domain and we classified AP2 transcription factors on the basis of the predicted amino acid sequences of the AP2 domains. From the NCBI Protein Database, 11 AP2 protein sequences from wheat and other plant species were retrieved and subjected to a multiple alignment analysis using ClustalW. A consensus sequence was obtained from the Clustal server and manually as well.

Phylogenetic analysis of AP2 protein

From the phylogenetic analysis, we can infer that wheat AP2 TF was found to be closely related to both *Papaver somniferum* and *Triticum aestivum*. The two branches get exchanged when different algorithms are used. *Papaver somniferum* is the closest member when Maximum Parsimony method is used. Although the branch-bootstrap value for *Papaver somniferum* being closer to *Triticum aestivum* is at 28, much higher than that of with its relatedness with *Brassica carinata* at 50. *Brassica carinata* and *Brassics juncea* are closely related in the maximum parsimony method with a bootstrap value of 99, a more accepted one. On the other hand, *Sisymbrium irio* has been shown to be closely related to *Capsella rubella* with a bootstrap value of 99 in minimum evolution method. *Arabidopsis lyrata subsp* has been shown to be closely related to *Arabidopsis arenosa* with a bootstrap method value of 73, *Arabidopsis arenosa and Olimarabidopsis pumila*. Rest all branches do not show much consensus due to the inefficiency of these character based methods when exact homology is not found in terms of sequence alignment.

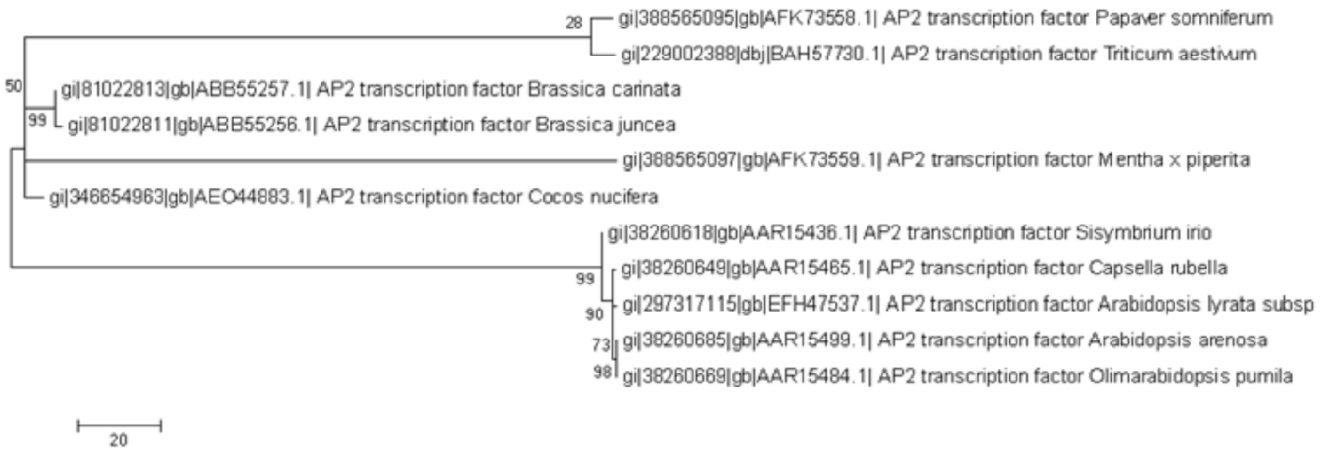

Figure 1: Phylogenetic tree constructed by Maximum likelihood and bootstrapping method

Structure Prediction and validation

A protein is functional when it folds into three dimensional structures that enclose its functional region. Prediction of protein structure offers insight into conformational properties and structure–function relationship [15]. We have modeled the consensus sequence obtained from the multiple sequence alignment of AP2 sequences. A protein structure is built for each of the 5 top-ranked alignments between the target sequence and the structures in the template library, Among the top 10 ranked alignments, the best template, PDB ID 2np0A, was used to model the protein. The final structure obtained with first rank had a C-score of 0.151 (C-score is the confidence score for high probability confirmation (Figure 2). C-score values range in between [-2 to +5]), a TM-score of 0.403 (TM score is a measure of global structural similarity between query and template protein), RMSD of 8.16 (RMSD is a measure of global structural alignment), IDEN value of 0.052 (IDEN is the percentage sequence identity), COV value of 0.669 (it represents the coverage of global structural alignment and is equal to the number of structurally aligned residues divided by length of the query protein.)

Further, the structure was validated using the PSVS server. The G-factor analysis for phi-psi angles and all dihedral angles were calculated for modeled residue. It was found that majority of amino acid residues of AP2 attains positive values on the G factor scale and thus signify its high probability conformation. These residues were found to be stable in terms of phi-psi angle and all dihedral angle G-factor values. In the Ramachandran Plot analysis (Figure 3), the structure showed a large number of residues falling in the allowed regions of the plot (83.0%). Also, a very small percentage of residues were found as outliers or in the disallowed regions of the plot (only 0.5%).

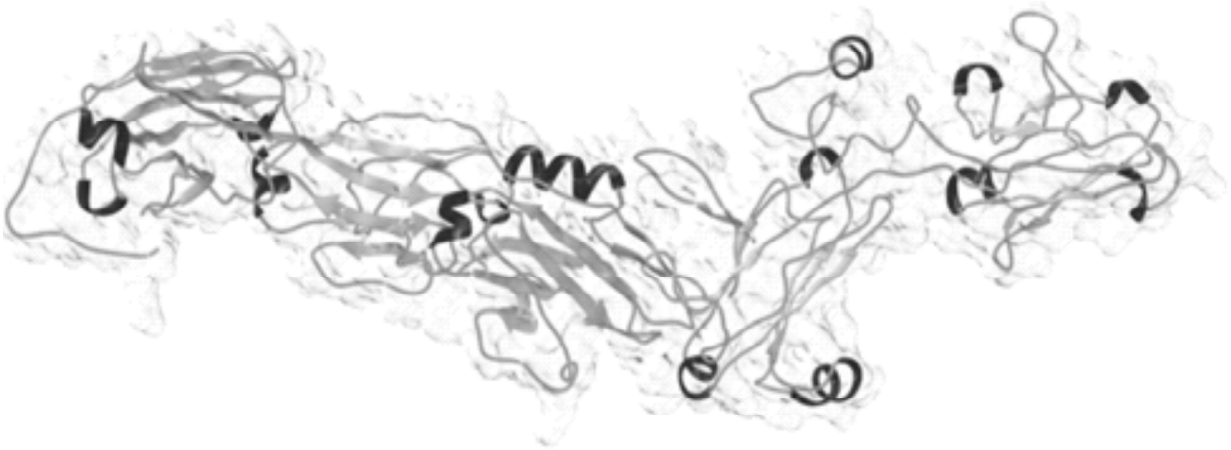

Figure 2: 3D structure model of AP2 produced using I-TASSER

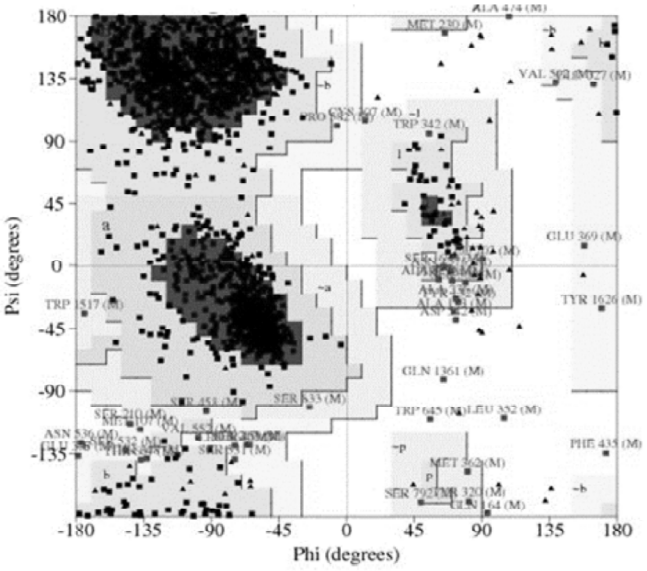

Figure 3: Ramchandran plots for each residue using ProCheck

In this diagram, the white areas correspond to conformations where atoms in the polypeptide come closer than the sum of their van der Waals radii. These regions are sterically disallowed i.e. 0.5% of all amino acids of this protein. The red regions correspond to conformations where there are no steric clashes, i.e. these are the allowed regions namely the á-helical and β-sheet conformations. Thus, the most favoured region has 83.0% of the total amino acid residues of modeled AP2 protein. The yellow areas show the allowed regions if slightly shorter van der Waals radii are used in the calculation, i.e. the atoms are allowed to come a little closer together. This brings out an additional region which corresponds to the left-handed β-helix i.e. 14.6% of amino acids fall in this region for AP2 protein.

CONCLUSION

The AP2/ERF transcription factor super family is one of the largest groups of transcription factors in plants, which includes at least one APETALA2 (AP2) domain. The present study involves *in silico* approaches, including secondary structure analysis, evolutionary trends and three dimensional protein analyses. The results revealed that *Brassics carinata* and *Brassica juncea* showed highest similarity in phylogenetic analysis. The protein models obtained showed the presence of a common super-secondary motif of sheet-helix-sheet type. The residues belonging to the same region consisting of the motif were found to have positive G-score value for phi-psi as well as all dihedral angles. The presence of conserved motif can be linked to its functionality in terms of binding to DNA. The protein models were validated with almost all residues falling in the allowed regions of Ramchandran plot, thus, establishing the goodness of predicted models. Computational analysis has been a time and cost effective approach and helps in aiding to experimental procedures. Any preliminary analysis can be supported with the use of *in silico* analysis. Thus, this study involves the use of computational techniques for AP2 protein model prediction which will pave way for a greater understanding towards its function in abiotic stress response and interaction with DNA. Such knowledge may also be useful in devising molecular methods to prevent damage of plants as a result of abiotic stresses.

ACKNOWLEDGEMENTS

Authors are thankful to the Department of Biotechnology, Govt. of India, for their financial support.

REFERENCES

Jones RA, Qualset CO, "Breeding crops for environmental stress tolerance", In: G.B. Collins and J.F. Petolino (eds.), 305-380, 1984.

Bray EAJ, Bailey SE, Weretilnyk E, "Responses to abiotic stresses, In: Biochemistry and Molecular Biology of Plants", American Society of Plant Physiologists, 1158-1176, 2000.

Boyer JS, "Plant productivity and environment", Science, 443-8, 1982.

Shinozaki K, Yamaguchi-Shinozaki K, "Molecular responses to dehydration and low temperature: differences and cross-talk between two stress signaling pathways", Current Opinion in Plant Biology, 217–223, 2003.

Blumwald E, "Sodium transport and salt tolerance in plants", Current Opinion in Cell Biology, 431–434, 2003.

Licausi F, Giorgi FM, Zenoni S, Osti F, Pezzotti M, Perata P, "Genomic and transcriptomic analysis of the AP2/ERF superfamily in *Vitis vinifera*", BMC Genomics, 719, 2010.

Licausi F, Ohme-Takagi M, Perata P, "APETALA2/Ethylene Responsive Factor (AP2/ERF) transcription factors: mediators of stress responses and developmental programs", New Phytology, 639-649, 2013.

De Boer K, Tilleman S, Pauwels L, Vanden Bossche R, De Sutter V, Vanderhaeghen R, Hilson P, Hamill JD, Goossens A, "APETALA2/ ETHYLENE RESPONSE FACTOR and basic helix–loop–helix tobacco transcription factors cooperatively mediate jasmonate-elicited nicotine biosynthesis", Plant Journal, 1053-1065, 2011.

www.ncbi.nlm.nih.gov/

www.ebi.ac.uk/Tools/msa/clustalw/

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S, "MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0", Molecular Biology and Evolution, 2725-2729, 2013.

Roy A, Kucukural A, Zhang Y, "I-TASSER: a unified platform for automated protein structure and function prediction", *Nature protocols*, 725-738, 2010.

psvs-1_5-dev.nesg.org/

Pettersen F, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE, "UCSF Chimera—a visualization system for exploratory research and analysis", *Journal of Computational Chemistry*, 1605-1612, 2004.

Kabsch W, Sander C, "Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features", Biopolymers, 2577-2637, 1983.