

This is a refereed journal and all articles are professionally screened and reviewed

**ORIGINAL ARTICLE****Identification of conserved domains and motifs for *TaWdhn13* gene in *Triticum aestivum* by *in silico* analysis****<sup>1</sup>Hassan Rasouli, <sup>1</sup>Danial Kahrizi, <sup>2</sup>Parivash Ghadernia**<sup>1</sup>*Agronomy and Plant Breeding and Biotechnology Department (APBB), College of Agriculture and Natural Resources, Razi University, Kermanshah, Iran.*<sup>2</sup>*Department of Microbiology in Brujerd Islamic Azad University*Hassan Rasouli, Danial Kahrizi, Parivash Ghadernia: Identification of conserved domains and motifs for *TaWdhn13* gene in *Triticum aestivum* by *in silico* analysis**ABSTRACT**

*TaWdhn13* gene has a crucial role as the cold-acclimation process in the wheat. *TaWdhn13* gene is transcriptionally activated and produces the accumulated proteins and metabolites and protection cell structure from freezing damage. The sequence of *TaWdhn13* gene (AC: AB297677) has been downloaded from NCBI database and *via* Conserved Domain Database (CDD) and DNA Motif Searching Database (DMSD), the conserved domains and motif has been recognized. The results showed *TaWdhn13* conserved domain includes: Dehydrin superfamily. Also, motifs structure for this gene includes: 2FE2S\_FER\_1 Motif, INTEGRIN\_BETA Motif, CTCK\_1 Motif, ANAPHYLATOXIN\_1 Motif, AGOUTI\_1 Motif, IGFBP\_N\_1 Motif, VWFC\_I Motif, EGF\_1 Motif and DEFENSIN Motif.

**Key words:** *TaWdhn13*, Conserved domain Data base (CDD), DNA Motif Searching Database (DMSD), *Triticum aestivum* L,

**Introduction**

Bioinformatics is an interdisciplinary research area, which may be defined as the interface between biological and computational sciences. It greatly helps in management of complex and scattered biological data, sequence analysis, algorithmic designing [7]. Domains, traditionally known as structurally independently folding units, are conserved functional units that may contain or more motifs. Domains are functional structural units. These units are conserved at the level of sequence and structure. Motifs are conserved across proteins at the level of sequence or structure or both. They include both short stretches of fixed residue length that act as sites for post-translational modification, phosphorylation, and longer sequence that form secondary structure for protein-DNA, protein-ion or protein-lipid interaction [7, 9]. *TaWdhn13* gene has a crucial role as the cold-acclimation process in the wheat. *TaWdhn13* gene is transcriptionally activated and produces the accumulated proteins and metabolites and protection cell structure from freezing damage [1]. In this study, we identify and characterize the conserved domain and motifs of *TaWdhn13* gene.

**Methods:****Download sequence:**

Download the complete sequence for *TaWdhn13* gene (with accession number: AB297677) in the NCBI database.

**Identification of the conserved domain:**

To identify conserved domain, we used the Conserved domain database (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>).

**Identification of the motifs in *TaWdhn13* gene:**

To analyze the nucleotide sequence we searched for the motifs and the motif search software (<http://www.genome.jp/tools/motif>) was used to identify the motifs in nucleotide sequence.

**Results and Discussion****Conserved domain:****Corresponding Author**

Hassan Rasouli, Agronomy and Plant Breeding and Biotechnology Department (APBB), College of Agriculture and Natural Resources, Razi University, Kermanshah, Iran.  
Tell: +989385436058  
E-mail address: hassanrasouli1978@yahoo.com

Our results showed *TaWdhn13* conserved domain includes: Dehydrin superfamily (Fig. 1).

#### Motif structure:

*In silico* analysis showed the motif structure for this gene includes (Table 1 and Fig. 2):

a) 2FE-2S\_FER\_1 Motif, or ferredoxins, iron-sulfur binding region signature. Motif 2Fe-2S is a structural motif, from the comparison of the coding proteins between rice and spinach chlorine monooxygenase (CMOs), rice CMO potentially shares two conservative motifs including a Rieske-type [2Fe-2S] [5] cluster and a mononuclear non-heme Fe binding sequence. These motifs are considered to be essential for the function of CMO [13].

b) INTEGRIN\_BETA Motif. Integrins are expressed on the cell surface as a noncovalently linked heterodimer consisting of  $\alpha$  and  $\beta$  subunit, which conveys specificity in cell-cell adhesion, cell-extracellular matrix (ECM) adhesion, immune cell recruitment, extravasation, and signaling [10].

c) CTCK\_1 Motif or C-terminal cystine knot signature. The cystine-knot motif, made up of three intertwined disulfide bridges, is a unique feature of several toxins, cyclotides and growth factors, and occurs in a variety of species, including fungi, insects, molluscs and mammals [4].

d) ANAPHYLATOXIN\_1 Motif. In the mammalian cell, Anaphylatoxins are able to trigger

j)

degranulation (release of substances) of endothelial cells, mast cells or phagocytes, which produce a local inflammatory response [17].

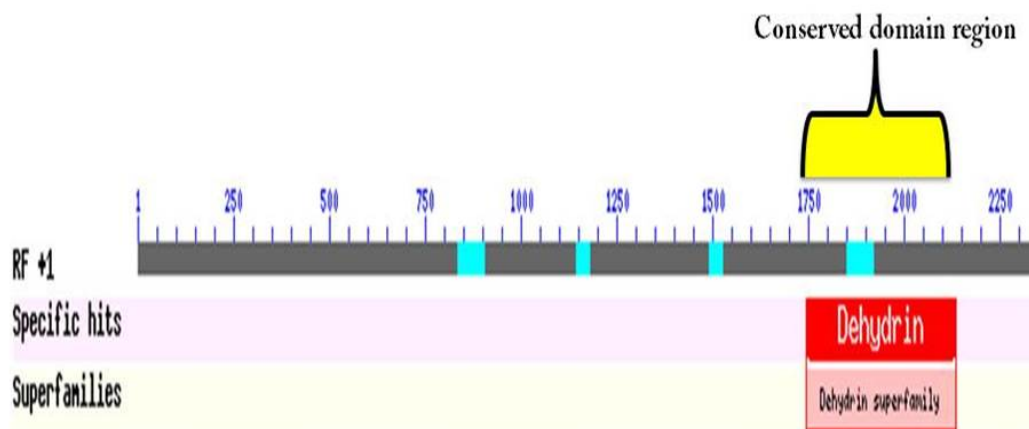
e) AGOUTI\_1 Motif. The AGOUTI\_1 Motif is a signaling motif and normally expressed in skin cells and interacts with the first melanocortin receptor (MCR1) to control pigmentation [9].

f) IGF1BP\_N\_1 Motif. Insulin-like growth factor (IGF) 1-binding proteins (IGFBPs) are a group of six homologous, yet distinct proteins (IGFBPs1–6) which bind both IGF-I and IGF-II with high affinity [16].

g) VWFC\_1 Motif. VWFC motif has conserved cysteine which was found in many cereal crops like rice [8].

h) EGF\_1 Motif. The EGF\_1 motif or EGF-like domain signature 1 is an evolutionary conserved protein domain, which derives its name from the epidermal growth factor where it was first described. It comprises about 30 to 40 amino-acid residues and has been found in a large number of mostly animal proteins [15].

i) DEFENSIN Motif. Defensins motifs are small Cysteine-rich cationic proteins found in both vertebrates and invertebrates. They have also been reported in plants. They are, and function as, host defense peptides. They are active against bacteria, fungi and many enveloped and nonenveloped viruses [18].



**Fig. 1:** Conserved domain for *TaWdhn13* gene in common wheat.








#### Discussion:

This computational analysis focuses on identifying a number of motifs structure and conserved domains in *TaWdhn13* gene. *In silico* analysis provides an efficient way to indicating motifs sequence and conserved domains in genome [11,12].

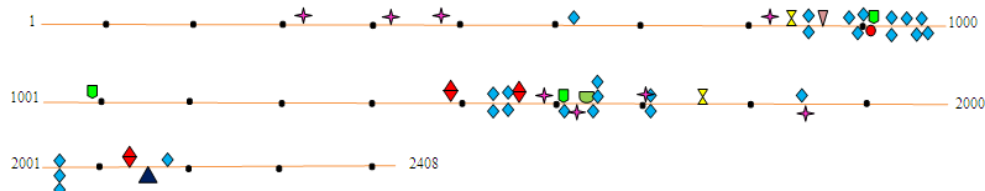
However, by using the *in silico* analysis we can predict and identify the motifs structure and conserved domain in gene sequence.

In this study a total of 9 motifs structure and one conserved domain identified in *TaWdhn13* gene sequence.

**Table 1:** Location and sequence to identified motifs in the *TaWdhm13* gene.

Motif	Motif location	Sequence	Symbol
2FE-2S_FER_1 Motif	1491..1499 1579..1587 2123..2131	CAGGTCGGC CTAGTCGGC CTGGACAGC	
INTEGRIN_BETA Motif	856..870 1771..1784	CCGCCGCCTCCTCCC CAGGGGCACGGCGC	
CTCK_1 Motif	1602..1640	CCACTCACAAGAGCACACGTGGTGCTTTCTCCACCCTCC	
ANAPHYLATOXIN_1 Motif	869..901 1063..1097 1565..1602	CCGCGCATCACGGCGCCGACGCGCCGCCAGCC CCGAGTGCACCGATCTCGGAGTGGATCGAGTAGCC CCCCACTTTATGAGTAGTCGGCAGTCACCTGCCACC	
AGOUI_1 Motif	836..878	CCTAGCGCTCTCACCACCGCCGCCCTCCTCCCAGCGGATC	
IGFBP_N_1 Motif	890..905	GCGCCGCCAGCCGCGC	
VWFC_1 Motif	564..616 805..859 806..859 841..891 843..897 844..901 848..905 859..915 860..915 862..915 863..915 874..928 901..945 907..945 910..968 911..968 922..975 923..975 925..975 1499..1546 1502..1546 1511..1568 1527..1570 1528..1570 1565..1607 1587..1640 1599..1643 1601..1643 1602..1643 1657..1708 1666..1708 1812..1860 1980..2034 1981..2034 1986..2034 2128..2175	CCACTCGGGAGCAGCTTGACGAGTTCTACCAACTTTCCAGACATCCAGCTC CCGACTCCCCCTTAGGAGCCAGAACCCTAAACCCTAGCGCTCTCACCACCGC CGC CGACTCCCCTTAGGAGCCAGAACCCTAAACCCTAGCGCTCTCACCACCGCC GC CGCCTCTCACCACCGCCGCCCTCTCCCAGCGATCACGGCGCCGACGC CCTCTCACCACCGCCGCCCTCTCCCAGCGATCACGGCGCCGACGCGCC GCC CTCTCACCACCGCCGCCCTCTCCCAGCGATCACGGCGCCGACGCGCCG CCAGCC CACCACCGCCGCCCTCTCCCAGCGATCACGGCGCCGACGCGCCGCCAG CCGCGC CCGCTCCTCCCAGCGATCACGGCGCCGACGCGCCGCCAGCCGCGTCAAC CTCTC CGCTCCTCCCAGCGATCACGGCGCCGACGCGCCGCCAGCCGCGTCAACC TCTC CCTCTCCCAGCGATCACGGCGCCGACGCGCCGCCAGCCGCGTCAACCTC TC CTCTCCCAGCGATCACGGCGCCGACGCGCCGCCAGCCGCGTCAACCTCT C CGATCACGGCGCCGACGCGCCGCCAGCCGCGTCAACCTCTACCAACCCAC CTC CGCGCTCAACCTCTCACCACCCACCTCTCATCCCTACAACCTAC CAACCTCTCACCACCCACCTCTCATCCCTACAACCTAC CCTCTCACCACCCACCTCTCATCCCTACAACCTACTGTTGAGGCCCGGTGAC CCTAGC CTCTCACCACCCACCTCTCATCCCTACAACCTACTGTTGAGGCCCGGTGACC CTAGC CCACTCTCATCCCTACAACCTACTGTTGAGGCCCGGTGACCCTAGCTCCTAC C CACCTCTCATCCCTACAACCTACTGTTGAGGCCCGGTGACCCTAGCTCCTACC CCTCTCATCCCTACAACCTACTGTTGAGGCCCGGTGACCCTAGCTCCTACC CAGCCACTCATGCGAGCACACGTGCGCCCTGCACTGCATGCCGCGC CCACTCATGCGAGCACACGTGCGCCCTGCACTGCATGCCGCGC CGAGCACACGTGCGCCCTGCACTGCATGCCGCGGATTCTGCCATCTAAC CACCC CCTGCACTGCATGCCGCGGATTCTGCCATCTAACCACCCAC CTGCACTGCATGCCGCGGATTCTGCCATCTAACCACCCAC CCCCACTTTATGAGTAGTCGGCAGTCACCTGCCACCCACTC CAGTCACCTGCCACCCACTCACAAGAGCACACGTGGTGCTTTCTCCACCCTC C CACCCTCACAAGAGCACACGTGGTGCTTTCTCCACCCTCCAAC CCACTCACAAGAGCACACGTGGTGCTTTCTCCACCCTCCAAC CCACTCACAAGAGCACACGTGGTGCTTTCTCCACCCTCCAAC CGCTCGTGCTGCACTTGCTTTACACAGCCACCTTCTCACAACCCAGGAC CTGCACTTGCTTTACACAGCCACCTTCTCACAACCCAGGAC CATCACGGAGAAGCTCCCCGTGGCCATGGTGATCACCAGCAGGCCACC CCAGCACCACTGGAATGAGCGGCTCGAAGACGCATGCCACCACAGCCAC CACC CAGCACCACTGGAATGAGCGGCTCGAAGACGCATGCCACCACAGCCACC ACC CACCCTGGAATGAGCGGCTCGAAGACGCATGCCACCACAGCCACCACC	

		CAGCACTAAGCCCAGCCGGTCTGCCACGCCCGCGCCCGACCCGCTAC	
EGF_1 Motif	344..355 410..421 461..472 811..822 1568..157 9 1615..162 6 1697..170 8 1846..185 7	CGCGCCCTGGCC CACCAGCGGTGC CACCAGTAGTGC CCCCTTAGGAGC CACTTTATGAGC CACACGTGGTGC CACACCAGGAC CACCAGCAGGCC	✦
DEFENSIN Motif	2131..215 9	CACTAAGCCCAGCCGGTCTGCCACGCC	▲



**Fig. 2:** Map of the motif structure sequence in the *TaWdhn13* gene.

#### Conclusion:

Our results showed in *TaWdhn13* gene, motifs structure includes: 2FE2S\_FER\_1 Motif, INTEGRIN\_BETA Motif, CTCK\_1 Motif, ANAPHYLATOXIN\_1 Motif, AGOUTI\_1 Motif, IGFBP\_N\_1 Motif, VWFC\_1 Motif, EGF\_1 Motif and DEFENSIN Motif. Also conserved domain in this gene include: Dehydrin superfamily.

#### References

- Kobayashi, F., M. Ishibashi, S. Takumi, 2012. Transcriptional activation *Cor/Lea* genes and increase stress tolerance through expression of a wheat *DREB2* homolog in transgenic tobacco. *Transgenic Research*, 17(5): 755-767.
- Zhengxin, C.H., 2012. Development of a Database Course for Bioinformatics. *Procedia Computer Science*, 9: 532-539.
- Mario, C., W. Rodrigo, S. Joakim, 2012. Biomedical and Bioinformatics Challenges to Computer Science. *Procedia Computer Science*, 1: 931-933.
- Shalini, L., K.R. Acharya, 2011. Tying the knot the cystine signature and molecular-recognition processes of the vascular endothelial growth factor family of angiogenic cytokines. *FEBS Journal*, 278: 4304-4322.
- Marwa, N.M.E., A.A. E-seuodi, A.A. Abdelhady, S.E.A. Edris, N.R. Abdelhamid, A.H.M. Hassan, 2011. Utilization of Some EST Derived SSR Markers to Detect Glutenin Gene Loci in Wheat. *Journal of Applied Sciences Research*, 7(12): 2147-2155.
- Ti-Cao, Z., Q. Qiao, Y. Zhong, 2012. Detecting adaptive evolution and functional divergence in aminocyclopropane-1-carboxylate synthase (ACS) gene family. *Computational Biology and chemistry*, 38: 10-16.
- Cannataro, M., R.W. dos Santos, J. Sundnes, 2009. Bioinformatics' challenges to computer science: Bioinformatics tools and biomedical modeling, in: G. Allen, J. Nabrzyski, E. Seidel, G. D. van Albada, J. Dongarra, P. M. A. Sloot (Eds.), ICCS (1), Vol. 5544 of Lecture Notes in Computer Science, Springer, pp: 807-809.
- Fan, C.H., Y.Z. Xing, H.L. Mao, T.T. Lu and B. Han, 2006. GS3, a major *QTL* for grain length and weight and minor *QTL* for grain width and thickness in rice, encodes a putative transmembrane protein. *Theor. Appl. Genet.*, 112: 1164-1171.
- McNulty, J.C., P.J. Jackson, D.A. Thompson, B. Chai, I. Gantz, G.S. Barsh, P.E. Dawson and G.L. Millhauser, 2005. Structures of the agouti signaling protein. *J Mol Biol.*, 346: 1059-1070.
- Adam, L.F., K. Heidi, C. Tresa, 2004. Cellular integrins function as entry receptors for human cytomegalovirus via a highly conserved disintegrin-like domain. *PANS*, 101: 15470-15475.
- Zhu, J.K., 2002. Salt and drought stress signal transduction in plants. *Annu. Rev. Plant. Biol.*, 53: 247-273.
- Lloyd, G., P. Landini, S. Bushby, 2001. Activation and repression of transcription initiation in bacteria. *Essays Biochem.*, 37: 17-31.
- Gray, J., P.S. Close, S.P. Briggs, G.S. Johal, 1997. A novel suppressor of cell death in plants encoded by the *Lls1* gene of maize. *Cell*, 89: 25-31.
- Rathinasabapathi, B., M. Burnet, B.L. Russell, D.A. Gage, P.C. Liao, G.J. Nye, P. Scott, J.H. Golbeck, A.D. Hanson, 1997. Choline monooxygenase, an unusual iron-sulfur enzyme catalyzing the first step of glycine betaine synthesis in plants: prosthetic group

- characterization and cDNA cloning. *Proc Natl Acad Sci USA.*, 94: 3454-3458.
15. Downing, A.K., V. Knott, J.M. Werner, C.M. Cardy, I.D. Campbell, P.A. Handford, 1996. Solution structure of a pair of calcium-binding epidermal growth factor-like domains: implications for the Marfan syndrome and other genetic disorders. *Cell.*, 85(4): 597-605.
  16. Rechler, M.M., 1995. In *Molecular Endocrinology: Basic Concepts and Clinical Correlations* (Weintraub, B. D., ed) pp. 155-180, Raven Press, Ltd., New York
  17. Hugli, T.E., 1986. *Biochemistry and Biology of Anaphylatoxins. Complement.*, 3(3): 111-27.
  18. Selsted, M.E., S.S. Harwing, T. Ganz, J.W. Schiling and R.I. Lehrer, 1985. Primary structure of three human neutrophil defensins. *J.Clin. Invest.*, 76: 1436-1439.