

Analysis of osmotin, a PR protein as metabolic modulator in plants

Malik Zainul Abdin^{1*}, Usha Kiran², Afshar Alam³

¹Department of Biotechnology, Faculty of Science, Jamia Hamdard, New Delhi-110062, India; ²Faculty of Engineering and Interdisciplinary Sciences, Jamia Hamdard, New Delhi-110062, India; ³Department of Computer Science, Jamia Hamdard, New Delhi-110062, India; Malik Zainul Abdin- Email: mزابدين@rediffmail.com; Tel: +91-11-26059688, Extn: 5583; *Corresponding Author.

Received November 01, 2010; Accepted December 04, 2010; Published January 22, 2011

Abstract:

Osmotin is an abundant cationic multifunctional protein discovered in cells of tobacco (*Nicotiana tabacum* L. var Wisconsin 38) adapted to an environment of low osmotic potential. Beside its role as osmoregulator, it provides plants protection from pathogens, hence also placed in the PRP family of proteins. The osmotin induced proline accumulation has been reported to confer tolerance against both biotic and abiotic stresses in plants including transgenic tomato and strawberry overexpressing osmotin gene. The exact mechanism of induction of proline by osmotin is however, not known till date. These observations have led us to hypothesize that osmotin could be regulating these plant responses through its involvement either as transcription factor, cell signal pathway modulator or both in plants. We have therefore, undertaken the present investigation to analyze the osmotin protein as transcription factor using bioinformatics tools. The results of available online DNA binding motif search programs revealed that osmotin does not contain DNA-binding motifs. The alignment results of osmotin protein with the protein sequence from DATF showed the homology in the range of 0-20%, suggesting that it might not contain a DNA binding motif. Further to find unique DNA-binding domain, the superimposition of osmotin 3D structure on modeled *Arabidopsis* transcription factors using Chimera also suggested absence of the same. However, evidence implicating osmotin in cell signaling were found during the study. With these results, we therefore, concluded that osmotin is not a transcription factor, but regulating plant responses to biotic and abiotic stresses through cell signaling.

Keywords: Osmotin, Osmoregulator, DNA binding motifs, Alignment.

Background:

The growth and productivity of crop plants depend largely on their vulnerability to environmental stresses. High salinity, water deficit, and temperature stress are the major constraints that limit agricultural production [1]. Plants respond to these conditions with an array of biochemical and physiological adaptations, which involve the function of many stress-related genes. Hence, any attempt to improve the stress tolerance requires a better understanding of the underlying physiological, biochemical and molecular events [2].

A common adaptation in plants is the production of compatible solutes and they accumulate to high levels without disturbing intracellular biochemistry [3]. Compatible solutes have the capacity to persevere the activity of enzymes that are in saline solutions. These compounds have minimal effect on pH or charge balance of the cytosol or luminal compartments of organelles. Compatible solutes include low molecular weight sugars, organic acids, polyols and nitrogen containing compounds such as amino acids, amides, imino acids, proteins and quaternary ammonium compounds [4].

Osmotin, a stress responsive multifunctional protein, has been reported to be involved in osmotolerance of the plants [5, 6]. Singh et al. (1985) [7] isolated this protein and its gene from tobacco var. Wisconsin 38 that was cultured for several generations on the culture medium with high NaCl concentration. It is a basic protein with a molecular weight of 24 KDa that accumulated in the tobacco cells and imparted to them osmotic stress adaptation. This function of osmotin is linked to its action as compatible

osmolyte that enhances osmotic potential of the cells. In addition to its role as osmoprotectant, osmotin is shown to exhibit antifungal activity against a broad range of fungal pathogens [8, 9].

Barthakur et al. (2001) [10] gave evidence that over expression of osmotin induces proline accumulation and confers tolerance to osmotic stress in transgenic tobacco. Biochemical analysis of tomato transformed with osmotin gene showed higher proline content in transformed plants under normal and stress conditions. 4 to 6 fold increase in proline content, over the wild type plants, was observed in leaves of strawberry plants transformed with osmotin gene even without subjecting them to the stress conditions [11]. It has been suggested that the synthesis of osmotin protein could induce synthesis and accumulation of certain solutes or could be involved in metabolic or structural changes of cellular proteins [12].

Based on the available literature and protein information on osmotin, we hypothesized that it may be involved in modulation of plant responses to biotic and abiotic stresses either directly by acting as transcriptional regulator for the genes encoding key enzymes of this pathway or as signaling molecule acting through intra cellular receptors. Being a 24 KDa protein, osmotin is within the range of low molecular weight transcription factors such as members of the GATA factor family of zinc finger transcription factors, WRKY transcription factors family etc. In the present study we therefore, evaluated osmotin protein for DNA binding domains and potential structural similarities with transcription factors available in databases.

Methodology:

Analysis of osmotin protein for DNA-binding domain:

Tobacco (*Nicotiana tabacum*) osmotin (1pcvA) (13) protein sequence and structural features were downloaded from RCSB-Protein Data Bank and UniProtKB/Swiss-Prot database for analysis. Amino acid sequence of osmotin protein in Fasta format was submitted to Compute pI/Mw at ExPASy Proteomics Server for calculating theoretical pI and molecular weight. The subcellular location of osmotin protein was predicted using TargetP 1.1 Server; <http://www.cbs.dtu.dk/services/TargetP> hosted at Technical University of Denmark, Denmark.

To identify the putative DNA binding sites on osmotin protein, amino acid sequence and structure were retrieved from NCBI databases and RCSB-Protein Data Bank, respectively. GT1 protein of Trihelix family of transcription factors from *Arabidopsis thaliana* was taken as control for this study. The 3D structure of GT1 (2jmwA) was downloaded from RCSB-Protein Data Bank. Osmotin and GT1 amino acid sequences and structures (as required) were submitted to DNA-binding motif prediction softwares; 2-zip server, DBD transcription factor prediction database, GYM 2.0 and Predictdnahth [14, 15, 16, 17].

Since these programs do not cover all types of DNA binding motifs especially those present in unique transcription factors, we also compared the osmotin protein sequence and structure with the transcription factors listed in Database of *Arabidopsis* transcription factors (DATF) [18] from

Arabidopsis. To find unique DNA binding domain on osmotin protein, homology studies were carried out using sequence and structure of transcription factor families listed in DATF. For this, amino acid sequence of individual *Arabidopsis* transcription factor was taken from the DATF. At least, amino acid sequence of one and at maximum three transcription factors were retrieved from each of 64 families. Since the crystal structure of these transcription factors were not available, their nucleotide sequences were submitted to NCBI to blast against PDB to get a homolog on which modeling of these transcription factors can be performed. The best match was saved and used for homology modeling of the protein structure of transcription factors. Amino acid sequence of the representative transcription factor was submitted to SWISS-MODEL workspace, automated mode, for homology modeling of protein's structure [19]. The structures were visualized and analyzed using Chimera. SuperPose program was used to study the superimposition of the structures of transcription factors with osmotin and GT.

Osmotin/ thaumatin like protein in Plants:

The gene and protein sequences of osmotin and osmotin/thaumatin-like proteins were retrieved from published completed genomes of *Arabidopsis lyrata*, *Arabidopsis thaliana*, *Carica papaya*, *Medicago truncatula*, *Glycine max*, *Cucumis sativus*, *Populus trichocarpa*, *Vitis vinifera*, *Oryza sativa*, *Zea mays* and *Sorghum bicolor*. Superfamily server was used for structural domain analysis for Osmotin/thaumatin-like proteins family [20].

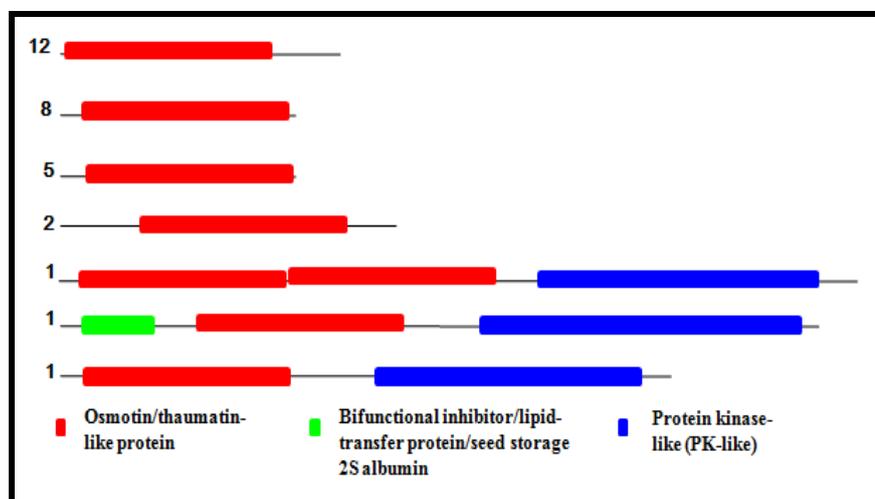


Figure 1: The domain architecture of Osmotin/thaumatin-like protein superfamily present in *Arabidopsis thaliana*

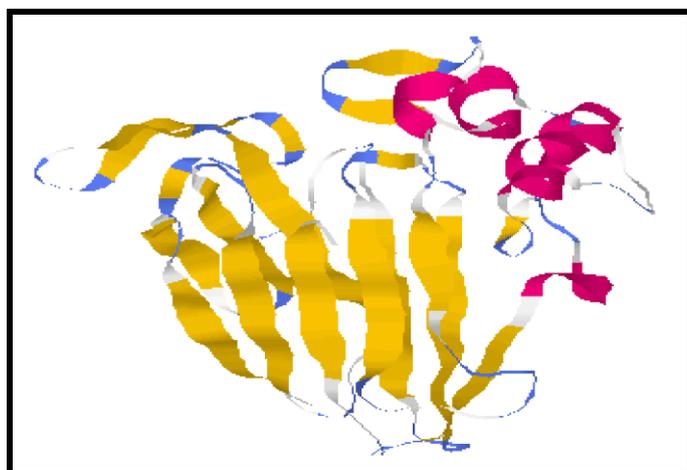


Figure 2: 3D structure of 1pcvA downloaded from RCSB PDB and visualised by Rasmol

Protein kinases ATP-binding region signature :		
		IGKGGFGTVYkGklpdasgrdiALK
AT4G18250.1	RYSFEKVKKMTNSFDHV	IGKGGFGTVYKGLPDASGRDIALKILKESKNGEEFINELVS 567
AT5G38280.1	RYSYTRVKKMTNSFAHV	LKGGGFGTVYKGLADS-GRDVAVKILKVSEGN EEGE FINEVAS 378
AT1G70250.1	RFYSYVQVKKMTKSFENV	LKGGGFGTVYKGLPDG-SRDVAVKILKESNEDGEDFINEIAS 506
Serine/Threonine protein kinases active-site signature		
		IvHFDIKpNILI
AT4G18250.1	GLEYLHNSCVSKI	IVHFDIKPQNILIDEDLCPKISDFGLAKLCKKKE SII SMLDARGTVGY 687
AT5G38280.1	GLEYLHNR CVTR	IVHFDIKPQNILMDENLCPKISDFGLAKLCKNKE SII SMLHMRGTFGY 498
AT1G70250.1	GLEYLHSHCVSR	IVHFDIKPQNILIDGDLCPKISDFGLAKLCKNNE SII SMLHARGTIGY 626

Figure 3: The Protein kinases signature domains present in Osmotin/thaumatin-like protein superfamily

Results and Discussion:

Evaluating osmotin protein using available bioinformatic tools, as direct/indirect metabolic modulator (transcription factor/cell signal molecule), will help in understanding the molecular adaptations by plants against abiotic and biotic stresses. Motif-discovery is often one of the first steps performed during computational analysis of gene-regulation. For any protein to act as a gene regulator (transcription factor), it must contain a DNA binding motif. The results of transcription factor prediction softwares showed the presence of HTH motif on GT1 protein thus, suggesting that it binds to DNA molecule. However, no DNA binding domains were detected on 3D structure of osmotin (**Table 1 see Supplementary material**). Thus, overall results suggest that osmotin protein lacks leucine zippers and HTH motifs, the two predominant and important DNA binding motifs. The alignment results of osmotin with the protein sequence from DATF showed the homology in the range of 0-20%, suggesting that it does not contain a DNA binding motif. In order to find unique domain, the superimposition of osmotin 3D structure was done on modeled Arabidopsis transcription factors using Chimera. The transcription factor, GT1 sequence was used as the control protein to check the workability of the program. GT1 completely superimposed on transcription factors belonging to Tri-helix family, whereas no significant superimposition was found with osmotin to any protein in the database. These results thus, indicate that there are no DNA binding motifs present in osmotin, similar to those represented by our test set (DATF).

Osmotin/ thaumatin like proteins in Plants:

Existing annotation and BLAST searches of well-characterized tobacco osmotin protein, used to identify osmotin/ thaumatin like protein homologs among the completed genome of *Arabidopsis lyrata*, *Arabidopsis thaliana*, *Carica papaya*, *Medicago truncatula*, *Glycine max*, *Cucumis sativus*, *Populus trichocarpa*, *Vitis vinifera*, *Oryza sativa*, *Zea mays* and *Sorghum bicolor* showed that it is present in all 11 plants, either alone or in combination of 5 different partner domains; Protein kinase-like (PK-like), Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin, Mitochondrial carrier, Class II aaRS ABD-related and ARM repeats (**Table 2 see Supplementary material**). Each partner domain is shown to occur in diverse sets of protein families. Nature often brings several domains together to form multidomain and multifunctional proteins with a vast number of possibilities [22]. Proteins with multiple domains have advantages, providing increased stability and new cooperative functions [22]. In a multidomain protein, each domain may fulfill its own function independently, or in a concerted manner with its neighbours. Thus, the modular nature of osmotin has suggested the multiple roles for it.

Structural domain assignments from the superfamily database showed that *A. thaliana* genome contains 31 proteins belonging to osmotin/thaumatin-like protein superfamily. Osmotin/thaumatin domain together with 2 different partner domains; PK-like and Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin occur in seven different domain combinations, among the members of this family (**Figure 1**). Protein kinases catalyze the phosphotransfer reaction fundamental to most

signaling and regulatory processes in the eukaryotic cell [23]. According to InterPro annotation, the three-dimensional fold of catalytic domain in

protein kinase is similar to domains found in proteins involved in cellular signalling including the catalytic domain of phosphoinositide-3-kinase (PI3K), the catalytic domain of the MHCK/EF2 kinase and choline kinase. They regulate a wide range of cellular process such as enzyme activation, assembly of macromolecules, protein localization and degradation. Three proteins in osmotin/thaumatin-like protein superfamily in *A. thaliana* contain this domain (AT4G18250.1, AT5G38280.1, AT1G70250) along with osmotin/thaumatin-like domain (**Figure 2 & 3**). The third domain belongs to plant lipid-transfer and hydrophobic proteins family and is mainly involved in lipid metabolism.

Responses to stress are generally complicated integrated circuits involving multiple pathways and cellular compartments [24]. The subcellular localization of a protein can provide important information about its function within the cell. The accurate function of proteins and their interaction networks relies greatly on the proper localization of each protein component. According to *in silico* estimates, *N. tabacum* osmotin is a cationic protein with theoretical pI of 8.13 and theoretical molecular weight of 26681.08 Da. The sequence alignment of *N. tabacum* osmotin (1pcvA) with *Arabidopsis* Osmotin/thaumatin-like protein superfamily shows that At4G 11650 (*atosm34*) is most similar to the *N. tabacum* osmotin, while AT4G 36000.1 is least. The score of location assignment program suggests *N. tabacum* osmotin to be a secretory protein i.e it is not bound to any membrane. According to the available evidences also, PR proteins are synthesized as precursors with an N-terminal signal peptide which mediates the transport. In general, the basic and neutral isoforms have been suggested to accumulate in vacuolar inclusion bodies [12]. Gau et al. (2004) [25] however, showed the accumulation of osmotin like protein in *Malus domestica* cv. Remo apoplast. Target P server results for 31 members of *A. thaliana* Osmotin/ thaumatin like protein superfamily suggests that all except seven (AT4G18250.1, AT1G20030.1, AT1G73620.1, AT1G77700.1, AT4G36000.1, AT4G38660.1 (M), AT4G38670.1 (C)) are secretory in nature. The secretory nature and multiple locations targeting of osmotin are in agreement to its multi functional role in plants exposed to biotic and abiotic stresses.

Apart from osmoregulation and functions involving partner domains, the osmotin could also act as modulator of transcription factors. This property of osmotin could be determined by the close protein-protein interactions with the transcription factors. Some *In silico* algorithms for studying protein-protein interactions have been formulated over the years [26]. These however, also suffer from limitations of either over dependence on sequence-based information or knowledge of the tertiary structure of both partners prior to predicting the quaternary structure. The question of which residues in the protein-protein interface actually interact often cannot be addressed and therefore, the exact function could be determined by the extensive studies involving labor-intensive experimental approaches. Consequently, experimental analyses are required to completely identify or assign a specific role for this protein.

The *in silico* analyses conducted in the present investigation indicate that osmotin/ thaumatin like protein superfamily is a large family of structurally similar proteins in the plants where osmotin/ thaumatin domain is present alone or with a partner domain. The specific function of this domain still needs to be addressed. The functions of partner domains however, are already determined. On this basis, we hypothesise that apart from osmoregulation, members of this super family are involved in diverse biological roles including cellular signal transduction.

Conclusion:

With increase in the number of completed genomes being published, determining protein function is one of the most important problems of this post-genomic era. Osmotin is a cationic protein with theoretical pI of 8.13 and theoretical molecular weight of 26681.08 Da. No DNA-binding motif was detected by DNA-binding motif prediction softwares; 2-zip server, DBD transcription factor prediction database, GYM 2.0, HTH motif prediction, Predictdnahth. Alignment results of osmotin protein with the protein sequence from DATF showed the homology in the range of 0-20%, suggesting that it does not contain a DNA binding motif. Further to find unique DNA-binding domain, the superimposition of osmotin 3D structure, done on modeled Arabidopsis transcription factors using Chimera also suggested absence of the same. With these results, we concluded that osmotin does not contain a DNA-binding motif and may be regulating plant responses to biotic and abiotic stresses by acting as a modulator of transcription factors or metabolic signaling. Further work is progress in our laboratory to elucidate the exact function of osmotin in plants under stress conditions.

Acknowledgement:

This study was supported by a research grant to Prof. M. Z. Abdin from UGC, Govt. of India. The Project Fellowship awarded to Usha Kiran by UGC is gratefully acknowledged.

References:

- [1] S Cherian *et al.* *Biologia Plantarum*, **50**: 481 (2006)
- [2] S Cherian & MP Reddy, *Biol. Plant.* **46**: 193 (2003)
- [3] HJ Bohnert *et al.* *Plant Cell*, **7**: 1099 (1995) [PMID: 12242400]
- [4] M Ashraf & PJC Harris. *Plant Sci.* **166**: 3 (2004).
- [5] NK Singh *et al.* *Plant Physiol.* **90**: 1096 (1989) [PMID: 16666857]
- [6] B Zhu *et al.* *Plant Mol. Biol.* **28**: 17 (1995) [PMID: 7787181]
- [7] NK Singh *et al.* *Plant Physiol.* **79**: 126 (1985) [PMID: 16664357]
- [8] DJ Yun *et al.* *Proc. Nat. Acad. Sci.* **94**: 7084 (1997)
- [9] LR Abad *et al.* *Plant Sci.* **118**: 11 (1996)
- [10] S Barthakur *et al.* *J. Plant Biochem. Biotechnol.* **10**: 31 (2001)
- [11] AM Husaini & MZ Abdin, *Plant Sci.* **174**: 446 (2008)
- [12] NK Singh *et al.* *Plant Physiol.* **85**: 529 (1987) [PMID: 16665731]
- [13] K Min *et al.* *Proteins* **54**: 170 (2004) [PMID: 14705035]
- [14] E Bornberg-Bauer *et al.* *Nucleic Acids Res.* **26**: 2740 (1998)
- [15] D Wilson *et al.* *Nucleic Acids Res.* **36**: D88 (2008) [PMID: 18073188]
- [16] C Bu Narasimhan *et al.* *J. of Computational Biol.* **9**:707 (2002)
- [17] WA McLaughlin & HM Berman, *J Mol Biol.* **27**:43 (2003) [PMID: 12818201]
- [18] A Guo *et al.* *Bioinformatics*, **21**: 2568 (2005)
- [19] K Arnold *et al.* *Bioinformatics*, **22**:195 (2006) [PMID: 16301204]
- [20] J Gough *et al.* *J Mol Biol.* **313**: 903 (2001) [PMID: 11697912]
- [21] C Chothia, *Nature* **357**: 543 (1992) [PMID: 1608464]
- [22] M Ostermeier & SJ Benkovic, *Adv. Protein. Chem.* **55**: 29 (2000) [PMID: 11050932]
- [23] RA Engh & D Bossemeyer, *Pharmacol Ther.* **93**: 99 (2002)
- [24] JE Dombrowski, *Plant Physiol.* **132**: 2098 (2003) [PMID: 12913164]
- [25] Achim E. Gau *et al.* *Eur J of Plant Pathology* **110**: 703 (2004)
- [26] I Ezkurdia *et al.* *Brief Bioinform.* **10**:233 (2009) [PMID: 19346321]

Edited by V Mathura

Citation: Abdin *et al.* Bioinformation 5(8): 336-340 (2011)

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.

Supplementary material:

Table 1: Results of DNA-binding motif prediction softwares on Osmotin protein

Software	Predicts	Submission	Results (Osmotin)	Results (GT 1)
2-zip server	Leucine zipper domain	Amino acid sequence	negative	negative
DBD transcription factor prediction database	Transcription factor	Amino acid sequence	negative	negative
GYM 2.0	HTH motif	Amino acid sequence	HTH motif found but with insignificant score	HTH motif found
predictdnath	HTH	PDB file	No DNA binding HTH motif was found	Helix of HTH motif found

Table 2: The domain architecture of Osmotin/thaumatin-like protein superfamily

Genome (Complete)	No.of Proteins in the superfamily	No. of Partner Domains	Domain Combination
Arabidopsis lyrata	30	2	1,2,3
Arabidopsis thaliana	30	2	1,2,3
Carica papaya	18	0	1
Medicago truncatula	18	0	1
Glycine max	57	1	1,4
Cucumis sativus	29	0	1
Populus trichocarpa	58	0	1
Vitis vinifera	29	0	1
Oryza sativa ssp. Japonica	44	1	1,2
Zea_mays subsp. mays	61	3	1,5,6
Sorghum bicolor	45	1	1,2

1. Osmotin
2. Protein kinase-like (PK-like)
3. Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin
4. Mitochondrial carrier
5. Class II aaRS ABD-related
6. ARM repeat

Table 3: Sequence alignment scores of members of *A. thaliana* Osmotin/ thaumatin like protein superfamily

SeqA	Name	Len(aa)	SeqB	Name	Len(aa)	Score
1	osmotin	246	2	AT1G18250.2	244	39
1	osmotin	246	3	AT1G19320.1	247	35
1	osmotin	246	4	AT1G20030.2	316	39
1	osmotin	246	5	AT1G70250.1	799	30
1	osmotin	246	6	AT1G73620.1	264	40
1	osmotin	246	7	AT1G75030.1	246	40
1	osmotin	246	8	AT1G75040.1	239	42
1	osmotin	246	9	AT1G75050.1	246	41
1	osmotin	246	10	AT1G75800.1	330	41
1	osmotin	246	11	AT1G77700.1	356	39
1	osmotin	246	12	AT2G17033.2	505	5
1	osmotin	246	13	AT2G17860.1	253	40
1	osmotin	246	14	AT2G24810.1	193	24
1	osmotin	246	15	AT2G28790.1	249	30
1	osmotin	246	16	AT4G11650.1	244	71
1	osmotin	246	17	AT4G18250.1	853	28
1	osmotin	246	18	AT4G24180.1	260	38
1	osmotin	246	19	AT4G36000.1	208	28
1	osmotin	246	20	AT4G36010.1	301	41
1	osmotin	246	21	AT4G38660.1	345	38
1	osmotin	246	22	AT4G38670.1	281	36
1	osmotin	246	23	AT5G02140.1	294	35
1	osmotin	246	24	AT5G24620.1	420	41
1	osmotin	246	25	AT5G38280.1	665	36