Computational predictions of common transcription factor binding sites on the genes of proline metabolism in plants

Usha Kiran¹ & Malik Zainul Abdin²*

¹Faculty of Interdisciplinary Research Studies, Jamia Hamdard, New Delhi-110062, India; ²Department of Biotechnology, Faculty of Science, Jamia Hamdard, New Delhi-110062, India; Malik Zainul Abdin – Email: mzabdin@rediffmail.com; Phone: +91-11-26059688, Extn: 5583; *Corresponding author

Received August 22, 2012; Accepted September 03, 2012; Published September 21, 2012

Abstract:
Proline, an imino acid, has been well documented to be associated with the stress response induced by abiotic factors such as drought, cold and salinity in plants and biotic factors such as bacterial and fungal attacks. However, the regulatory mechanisms controlling proline metabolism, intercellular and intracellular transport and connections of proline to other metabolic pathways are poorly understood. F-MATCH analysis combined with composite module analysis (CMA) revealed that the binding sites matching matrices for O2 and OCSBF1 were overrepresented in the promoters of differentially expressed proline metabolism genes. The presence of MYBAS1 consensus binding sites occurring in combination with O2 and OCSBF1 in the promoters of genes of proline biosynthesis pathway and SBF1 and GT1 consensus binding sites occurring in combination with O2 and OCSBF1 in the promoters of proline catabolic pathway genes suggest their involvement in modulation of proline metabolism and its accumulation in plants.

Keywords: Proline, Stress, Composite module analysis, Promoter, Transcription factor binding sites

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, 2]. Plants respond to these conditions with an array of biochemical and physiological adaptations, which involve the function of many stress-related genes. Proline accumulation has been well documented during drought [3], high salinity [4], high light and UV irradiation [5], heavy metals [6], oxidative stress [7] and in response to biotic stresses [8, 9]. Several functions are proposed for the accumulation of proline in tissues exposed to stresses. These include abiotic and biotic osmotic adjustments, C and N reserve for growth after stress relief, detoxification of excess ammonia, stabilization of proteins and/or membranes and scavenger of free radicals [7, 8, 10].

Accumulation of proline in plant cells under stress environment could be due to either its de novo synthesis or decreased degradation or both. Synthesis of proline in plants occurs in the cytosol and in the plastids (like chloroplasts in green tissues) and involves the sequential action of pyrroline-5-carboxylate synthetase (P5CS) and pyrroline-5-carboxylate reductase (P5CR), which convert glutamate to pyrroline-5-carboxylate (P5C) and P5C to proline, respectively [11, 12]. Several studies have indicated that P5CS is the critical enzyme in proline biosynthesis under salt and water stress [13, 14]. Proline biosynthesis is controlled by the activity of two P5CS genes in Arabidopsis, encoding one housekeeping and one stress-specific P5CS isoform. Arabidopsis P5CS1 is induced by osmotic and salt stresses and is activated by H2O2-derived signals and abscisic acid (ABA)-dependent pathway [4, 15]. P5CR is encoded by only one gene but the enzyme seems to be active in chloroplasts and cytosol [16, 12].
BIOINFORMATION

Proline is oxidized by sequential action of proline dehydrogenase (PDH), which converts proline to P5C and D1-pyrroline-5-carboxylate dehydrogenase (PSCDH), which converts PSC to glutamate [17, 18]. The oxidation of proline generates NADP/ NADPH cycling or redox balance therefore, important for the cell. The enzyme PDH is bound to the inner membrane of mitochondria. An alternative source of substrate for the PCDH enzyme can be derived from the conversion of arginine to ornithine and subsequent catabolism to PSC by ornithine aminotransferase [11]. PSCDH is a single copy gene in Arabidopsis and the encoded protein is mitochondria localized [10].

Whereas proline biosynthesis is upregulated by light and osmotic stresses, proline catabolism is activated in the dark and during stress relief [19, 20]. PSCS1 gene activation and proline accumulation is promoted by light and repressed by brassinosteroids [21, 19]. Under non-stressed conditions, phospholipase D (PLD) functions as a negative regulator of proline accumulation [22]. On the other hand, calcium signaling and phospholipase C (PLC) trigger P5CS transcription and proline accumulation during salt stress. However, in some halophyte PLD functions as positive regulator, whereas PLC exerts a negative control on proline accumulation [23, 24]. Calcium signals can be transmitted by a specific CaM4 calmodulin, which interacts with the MYB2 transcription factor [25]. Conversion of PSC to proline is not a rate-limiting step in proline biosynthesis, yet the control of PSCR activity implies a complex regulation of transcription, which was shown to be under developmental and osmotic regulation [12]. Promoter analysis of Arabidopsis PSCR identified a 69-bp promoter region that is responsible for tissue-specific expression [26]. However, trans-acting factors that can bind to this promoter region have not yet been identified.

Downregulation of PDH expression during stress is widely accepted as one control point that can promote proline accumulation under stress [17, 27, 28]. PDH transcription is activated by rehydration and proline, but repressed by dehydration; thus, preventing proline degradation during abiotic stress [29]. Promoter analysis of PDH1 identified the proline and hypo-osmolality-responsive element (PRE) motif A1CTCA, which is necessary for the activation of the PDH gene [30]. Basic leucine zipper protein (bZIP) transcription factors (AtbZIP3-2, -11, -44, -53) have been identified as candidates for binding to this motif [31]. The PSCDH gene is expressed at a low basal level in all Arabidopsis tissues, and can be upregulated by proline [32]. A short sequence similar to the PRE motif has been identified on the promoters of PSCDH genes in Arabidopsis and cereals [33].

Although the importance of proline accumulation conferring hyposalinity stress tolerance has been demonstrated well, the regulatory molecules as well as the molecular signals involved in the expression of proline biosynthetic genes are not understood. Comprehensive studies are required at physiological, molecular and genetic levels to explore the signal transduction events of proline synthesis and degradation. An insight into proline metabolism would be of interest to both those seeking to better understand plant stress physiology as well as those seeking to understand metabolic regulation. In the present study, we aim to analyze promoters of genes involved in proline metabolism in order to improve the understanding of the underlying physiological, biochemical and molecular events in stress tolerance by the plant.

Methodology:

Genes involved in proline biosynthesis and osmoregulation
Sequences of seven genes (Delta1-pyrroline-5-carboxylate synthetase 1, Pyrroline-5-carboxylate reductase, Proline oxidase, Putative / osmotic stress-responsive proline dehydrogenase, Delta-pyrroline-5-carboxylate dehydrogenase, Proline transporter 1, Proline transporter 2, Proline transporter3) of proline biosynthesis in Arabidopsis thaliana were retrieved after mapping the Locus IDs to the ATIGRS gene identifiers Table 1 (see supplementary material).

Promoter sequence analysis of differentially expressed genes of proline biosynthesis
Promoters sequence analysis of differentially expressed genes of proline metabolism was done using BIOBASE Knowledge Library Plant Edition (BKL-Plant) and ExpPlain Plant Analysis system. The promoter window of -1,000 to +100bp was uploaded into ExpPlain PlantAnalysis System, F-Match module was used to identify transcription factor binding sites overrepresented in differentially expressed proline biosynthesis gene sets against a background set of 200 ubiquitously present genes. The Composite Module Analysis was then used to determine which combination of binding sites, or Composite Module, was most commonly found within the sets of genes. Matrices with a Yes/ No score >1.3, pvalue <0.05, and Matched promoters p-value <.01 from F-Match analyses were selected for the composite module analysis (CMA).

Methodology:

Results and Discussion:

Regulation of gene expression plays an important role in a variety of biological processes such as development and responses to environmental stimuli including biotic and abiotic stresses. These responses are modulated by transcriptional
regulation of various genes. One of such plant responses to environmental stresses is the accumulation of proline. Although proline metabolism has been studied for very long time in plants, little is known about the signaling pathways involved in its regulation. Proline biosynthesis is activated and its catabolism repressed during dehydration, whereas rehydration triggers the opposite regulation [13, 34]. Based on the literature and genome information, Arabidopsis genome was parsed for the proline biosynthetic genes by using mapviewer. In general, cellular concentration of compatible solutes can be regulated by increasing biosynthesis, decreasing degradation, and/or modifying rates of uptake or release of these compounds. Therefore, the genes were divided into three sub-groups according to the metabolic functions. Group 1 of pyrroline-5-carboxylate synthetase (P5CS) and pyrroline-5-carboxylate reductase (P5CR). There are two P5CS gene loci in the nuclear genome of Arabidopsis thaliana. Out of these, AtP5CS1 is reported to be involved in stress tolerance. Group 2 of osmotic stress-responsive proline dehydrogenase (PO/ PDH) and pyrroline-5-carboxylate dehydrogenase (P5CDH). Group 3 of proline transporter with affinity for glycine betaine, proline and GABA (AtProT1), proline transporter 2 (AtProT2) and Proline transporter 3 (AtProT3) Table 1.

| Gene Symbol | TransPos | Match Display | FBS | Overrepresented |
|-------------|----------|---------------|-----|----------------|
| P5CS1       | AT3g2131 | -10 -10       | 1   | 1              |
| P5CS2       | AT1g014266 | -10 -10       | 2   | 2              |
| P5CR        | AT1g014266 | -10 -10       | 2   | 2              |
| P5CDH       | AT1g014266 | -10 -10       | 2   | 2              |
| ProT1       | AT3g014266 | -10 -10       | 2   | 2              |
| ProT2       | AT3g014266 | -10 -10       | 2   | 2              |

Figure 2: Over represented Transcription factor binding sites (TFBS) in Proline biosynthesis promoter set using F-Match Analysis.

Composite Module Analysis of proline biosynthetic Genes

A large number (>1,500) of transcription factors (TFs) in plants, control the expression of tens or hundreds of target genes in various, sometimes intertwined, signal transduction cascades. Transcription factor binding sites (TFBSs) are the functional elements that determine the timing and location of transcriptional activity. In plants and other higher eukaryotes, these elements are primarily located in the long non-coding sequences upstream of a gene, although functional elements in introns and untranslated regions have been described as well. The discovery of regulatory motifs and their organization in promoter sequences is an important first step to improve understanding of gene expression and regulation. Since coexpressed genes are likely to be regulated by the same TF, the identification of shared and thus overrepresented motifs insets of potentially co-regulated genes may provide an insight to the regulation of expression of whole metabolic pathway.

F-Match analysis compares the number of sites found in a query sequence set against the background set and provides, as results, the Position Weight Matrices (PWMs) whose frequencies are higher in the query sequence set compared to the background set. The F-match analysis results showed over represented TFBS in the proline biosynthetic gene set (Figure 1). The composite module analysis was used to determine the combination of binding sites most commonly found within the sets of genes. Upon analysis, the binding sites matching matrices for Opaque 2 (O2) and Ocs element binding factor (OCSBF-1) appeared most commonly in differing combinations within the promoters of proline biosynthesis genes (Figure 2 and Table 2 (see supplementary material).

The presence of O2 and OCSBF-1 consensus binding sites occurring in combination within the promoters of proline biosynthetic genes suggests a role for bZIP family proteins in the regulation of expression of this gene set. These proteins are of particular interest because they represent a group of plant transcriptional regulators that are similar to Fos [35] and CREB [36] families of proteins which act by way of conserved DNA binding domains and diverged effect or domains [37]. The information available on O2 and closely related monocot genes indicates that they regulate seed storage protein production by interacting with the PBF protein. The data derived from monocot and dicot species suggest that homologues of group S bZIPs are transcriptionally activated after stress treatments [38]. The presence of O2 and OCSBF-1 consensus binding sites in the members of proline metabolism genes set suggests common sensor for the concerted regulatory control of proline metabolism. Similar regulatory control of proline metabolism is seen upon illumination, where proline biosynthesis upregulated by light and osmotic stresses while proline catabolism is downregulated [19-21]. The blast analysis showed O2 and OCSBF1 show homology to the members of C and S groups of Arabidopsis AtbZIP transcription factor family, respectively. Blast results showed transcription factors closest to O2 are AtbZIP10 and AtbZIP25 and to OCSBF1 are AtbZIP53 and AtbZIP2.

Weltmeier et al. [39] used chromatin immunoprecipitation (ChIP) assay to show that ProDH is a direct target of the group S bZIP transcription factor AtbZIP53 and group C AtbZIP10. Promoter analysis of Arabidopsis P5CR identified a 69-bp promoter region that is responsible for tissue-specific expression [38]. The trans-acting factors that can bind to this promoter region, however, have not yet been identified. The regulation of other genes in proline metabolism by specific transcription factor is yet to be studied.

Composite module analysis also showed consensus binding sites of MYBA SL in P5CSI and P5CR, SBF-1 in PDH and GT-1 in P5CDH promoters (Supplementary data) suggesting that the proline metabolism gene cluster is likely to be under the combinatorial control of more than one class of transcription factors. Plant MYB proteins have been shown to regulate diverse developmental processes, as well as being involved in environmental signaling and secondary metabolism [40]. SBF-1 (SET binding factor 1) is a pseudo-phosphatase related to the myotubularin family of dual specificity phosphatases and binds to the silencer region of a chalcone synthase promoter [41]. The consensus sequence resembles the binding site for the GT-1
factor in light-responsive elements of the pea rbcS-3A gene suggesting that they may have similar functions [41]. Plant transcription factor GT-1 was identified by its specific binding activity to Box II, a promoter cis-element with the core DNA sequence 5-GGTAAA found initially in light-regulated genes and may activate transcription through direct interaction with the transcriptional pre-initiation complex [42].

The F-Match and CM analyses with the published literature suggest involvement of both trans- and cis-acting regulatory elements that sense proline levels either directly or indirectly. Therefore, to determine the precise regulation mechanism of proline metabolism, reverse genetic approaches (random or inducible counterparts) in combination with microarray technology may be done. Such studies with the metabolomic analyses should pave the way towards a better understanding of the functional diversification of plant TFs and regulatory molecule under stress conditions.

Conclusion:
The composite module analysis was used to determine the combination of binding sites most commonly found within the sets of proline metabolism genes. Upon analysis, the binding sites matching matrices for O2 and OCSBF-1 appeared most commonly in differing combinations within the promoters of proline biosynthesis genes. O2 and OCSBF1 showed homology with the members of C and S groups of AtbZIP transcription factor family, respectively. The presence of MYBAS1 consensus binding sites in combination with O2 and OCSBF1 in the promoters of genes of proline catabolic pathway genes suggest involvement of these transcription factors in the regulation of cellular levels of proline in plants exposed to abiotic and biotic stresses.

Acknowledgement:
Authors acknowledge UGC, Ministry of Human Resource Development, Government of India for the financial assistance and UGC-SAP, Government of India for providing the infrastructure for doing the computational work.

References:
[1] Sahi C et al. Funct Integr Genomics. 2006 6: 263 [PMID: 16819623]
[2] Cherian S et al. Biologia Plantarum. 2006 50: 481
[3] Choudhary NL et al. Indian J Biochem Biophys. 2005 42: 366 [PMID: 16955737]
[4] Yoshihara H et al. Plant J. 1995 7: 751 [PMID: 7773306]
[5] Saradhi PP et al. Biochem Biophys Res Commun. 1995 6: 209 [PMID: 7726821]
[6] Schat H et al. Physiol Plant. 1997 101: 477
[7] Yang SL et al. J Plant Physiol. 2009 166: 1694 [PMID: 19446917]
[8] Fabro G et al. Mol Plant Microbe Interact. 2004 17: 343 [PMID: 15077666]
[9] Haudecoeur E et al. Proc Natl Acad Sci. 2009 106: 14587 [PMID: 19706545]
[10] Smirnoff N et al. Phytochemistry. 1989 28: 1057
[11] Delauney et al. The Plant Journal. 1993 4: 215
[12] Verbruggen et al. Plant Physiol. 1993 103: 771 [PMID: 8022935]
[13] Kishor P et al. Plant Physiol. 1995 108: 1387 [PMID: 12228549]
[14] Yamada et al. J Exp Bot. 2005 56: 1975 [PMID: 15928013]
[15] Verslues PE et al. Plant Mol Biol. 2007 64: 205 [PMID: 17318317]
[16] Rayapati P et al. Plant Physiol. 1989 91: 581 [PMID: 16667072]
[17] Verbruggen N et al. Plant Physiol. 1993 103: 771 [PMID: 8022935]
[18] Deuschle K et al. Plant J. 2001 27: 345 [PMID: 11532180]
[19] Hayashi F et al. Plant Cell Physiol. 2000 41: 1096 [PMID: 11148267]
[20] Xue X et al. BM B Rep. 2009 42: 28 [PMID: 19192390]
[21] Abraham E et al. Plant Mol Biol. 51: 363 2003 [PMID: 12602867]
[22] Thierry L et al. J Biol Chem. 2004 279: 14812 [PMID: 14742440]
[23] Knight H et al. Plant J. 1997 12: 1067 [PMID: 9410048]
[24] Parre E et al. Plant Physiol. 2007 144: 503 [PMID: 17369432]
[25] Yoo JH et al. J Biol Chem. 2005 280: 3697 [PMID: 15569682]
[26] Hua X et al. FEBS Lett. 1999 458: 193 [PMID: 10481064]
[27] Verslues PE et al. Plant Mol Biol. 2007 64: 205 [PMID: 17318317]
[28] Sharma S and Verslues PE, Plant Cell Environ. 2010 33: 1838 [PMID: 20545884]
[29] Kiyosue T et al. Plant Cell. 1996 8: 1323 [PMID: 8776899]
[30] Satoh R et al. Plant Physiol. 2002 130: 709 [PMID: 12376638]
[31] Satoh R et al. Plant Cell Physiol. 2004 45: 309 [PMID: 15047879]
[32] Deuschle K et al. Plant J. 2001 27: 345 [PMID: 11532180]
[33] Ayliffe MA et al. Mol Gen. Genomics. 2005 274: 494 [PMID: 16179990]
[34] Lehnmann S et al. Amino Acids. 2010 39: 949 [PMID: 20204435]
[35] Van Beveren C et al. Cell. 1983 32: 1241 [PMID: 6301687]
[36] Gonzales GA et al. Cell. 1989 59: 675 [PMID: 2573431]
[37] Jakoby MB et al. Trends Plant Sci. 2002 7: 106 [PMID: 11906833]
[38] Hua X et al. FEBS. 1999 458: 193 [PMID: 10481064]
[39] Weltmeier F et al. EMBO J. 2006 25: 3133 [PMID: 16810321]
[40] Jin H & Martin C, Plant Mol Biol. 1999 41: 577 [PMID: 10645718]
[41] Lawton MA et al. Plant Mol Biol. 1991 16: 235 [PMID: 1893909]
[42] Le Gourrierec et al. Plant J. 1999 18: 663 [PMID: 10417717]

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

Citation: Kiran & Abdin, Bioinformation 8(18): 886-890 (2012)

Edited by P Kangueane
Supplementary material:

Table 1: Genes involved in proline metabolism and transport

| Group | Gene symbol | Description | Accession       | TSS            |
|-------|-------------|-------------|-----------------|----------------|
| I     | P5CS1       | Delta1-pyrroline-5-carboxylate synthetase 1 | At2g39800.1     | Chr.2.16610040 (-) |
|       | P5CR        | Pyrroline-5-carboxylate reductase | At5g14800       | Chr.5.4787842 (-) |
| II    | PDH(PO)     | Proline oxidase, putative/ osmotic stress-responsive proline dehydrogenase | At5g38710       | Chr.5.15518475 (+) |
|       | P5CDH(ALDH12A1) | Mitochondrial Delta-pyrroline-5-carboxylate dehydrogenase | At5g62530       | Chr.5.25120524 (-) |
| III   | ProT1       | Proline transporter with affinity for glycine betaine, proline and GABA | At2g99890       | Chr.2.16662731 (+) |
|       | ProT2       | Proline transporter 2 | At3g55740       | Chr.3.20706491 (+) |
|       | ProT3       | Proline transport3 | At2g36590       | Chr.2.15352416 (-) |

Table 2: Composite Module Analysis of proline biosynthesis pathway genes

| Gene symbol | TRANSPRO ID | PSOCBSBL_01 | PSQ2 | PSQ2_Q1 |
|-------------|-------------|-------------|------|---------|
| P5CS1       | ATH_3213    | "-286 to -282" | "-250 to -240" | "-250 to -240" |
| P5CR        | ATH_14248   | "-854 to -850" | "-834 to -774" | "-834 to -774" |
| PO          | ATH_16278   | "-825 to -821" | "-792 to -780" | "-792 to -780" |
| ALDH12A1    | ATH_18736   | "-471 to -467" | "-533 to -465" | "-533 to -465" |
| ProT1       | ATH_3223    | "-636 to -632" | "-756 to -712" | "-756 to -712" |
| ProT2       | ATH_8478    | "-914 to -910" | "-951 to -941" | "-951 to -941" |
| ProT3       | ATH_2891    | "-351 to -347" | "-353 to -308" | "-353 to -308" |