Molecular Characterization and Phylogenetic Analysis of BZIP Protein in Plants

Dhivya Selvaraj1, Arul Loganathan2 and Sathishkumar Ramalingam1*

1Molecular Biology Laboratory, Department of Biotechnology, Bharathiar University, Coimbatore -46, India
2Centre for Plant Molecular Biology, Tamilnadu Agriculture University, Coimbatore-46, India

Abstract

BZIP are a class of dimeric sequence specific DNA-binding proteins, is bipartite in structure containing region of enriched basic amino acids which is adjacent to leucine zippers. It is characterized by several leucine residues regularly spaced at seven amino acid intervals, basic region directly contacts with DNA. The leucine zipper mediates heterodimerization and homodimerization of protein monomers through parallel interactions which is unique to eukaryotes. The plant Arabidopsis thaliana genome shows 67 BZIP proteins. We have predicted dimeric properties of alpha helical leucine zipper and coiled coil structure of BZIP proteins in plants. In this analysis the length of leucine zippers, placement of asparagines in the hydrophobic interface and presence of interhelical electrostatic interactions were focused. Phylogenetic tree was also constructed by studying evolutionary relationship of BZIP existing among the plants.

Keywords: BZIP; Transcription factor; Dimerization; Leucine zippers; Biophysical properties; Phylogenetic relationship

Introduction

Growth and development of all organisms depends on respective gene expression which is mainly controlled by transcription factors. Transcriptional regulators can be grouped into families of related proteins (Michel et al., 2001). The basic leucine zipper (BZIP) is one among the transcriptional regulatory factors that have been conserved in all eukaryotes. BZIP protein has DNA binding domain consisting of rich regions of basic amino acids that binds to DNA and so called leucine zippers. It consists of several heptad repeats of hydrophobic residues which cause dimerization. BZIP basic region shows a high degree of sequence similarity with Homo sapiens and Arabidopsis thaliana and contain two invariant residues of Asparagine and Arginine.

BZIP–DNA complex consists of two α helices lying perpendicular to the DNA, associated in a coiled coil structure with basic region contacting a half site in the DNA major groove. The previous study reveals that BZIP structures shows functional variability of conserved residues in DNA recognition (Maria et al., 2003). The genome of Arabidopsis thaliana have been sequenced and annotated. Findings suggest that BZIP proteins are important for pathogen defense, light- induced signaling, seed maturation and flower development in plants (Christopher et al., 2004). BZIP proteins form homodimers and heterodimers depending on the amino acid sequence of the leucine zipper (O’Shea et al., 1992). In the previous study of basic region of EmBp-1, eight or ten conserved residues were found in other leucine zipper proteins (Guiltinan et al., 1990).

Leucine zippers of monomeric BZIP have structural repeats of two α helical turns and the repeat is termed as heptad, with each seven positions assigned as a,b,c,d,e,f and g. The positions d, e and g are near to leucine zipper interface and shows dimerization specificity. Amino acids in a and d positions are hydrophobic that lies on same side of the helix. These hydrophobic amino acids interact interhelically with hydrophobic amino acids in the same a and d positions of the second α-helix of the leucine zipper that stabilize the dimerization property of the protein. The Amino acid leucine has better stabilizing property than other amino acids (Michel et al., 2001) Dimerization specificity is regulated by amino acids in the a, e and g positions. The charged amino acids present in g and e positions actively involve in the formation of attractive electrostatic interhelical interactions. These interactions are denoted as g-e-a’ where the prime (’ ) indicates a residue on the second α helix of the dimeric leucine zipper. Oppositely charged amino acid interactions promote dimerization specificity where as similarly charged amino acids shows repulsion and thereby inhibiting homodimerization. BZIP plays an important role in Abscisic Acid (ABA) signaling pathways in Arabidopsis. Through quantitative RT-PCR, it is analyzed that most of OsbZIPS were induced by ABA, ACC and abiotic stress. The RT-PCR reveals that rice BZIP has a positive role in drought tolerance (Lu et al., 2009). Phylogenetic analysis of BZIP protein was done in algae, mosses, ferns, gymnosperms and angiosperms. The result suggests that the ancestor of green plants possess four bZIP genes that actively involved in oxidative stress and also in light-dependent regulations. (Luiz et al., 2008).

In this study, 109 motifs of BZIP in the genome of O. sativa japonica were analyzed by comparing with the model plant Arabidopsis thaliana and other plants. None of the proteins are homologous to animal BZIP proteins but they have similar amino acids to regulate dimerization specificity. BZIP protein sequences from different plant source like Phaseolus vulgaris, Capsicum annuum, Nicotiana tabacum, Antirrhinum majus, Hyaenithus orientalis, Malus x domestica, Phaseolus vulgaris, Triticum aestivum, Vitis vinifera, Glycine max, Lycopersicon esculentum, Catharanthus roseus, Spinacia oleracea and Psophocarpus tetragonolobus have been annotated. This analysis

*Corresponding author: Sathishkumar R, Molecular Biology Laboratory, Department of Biotechnology, Bharathiar University, Coimbatore-46, India, E-mail: rsathish@buc.edu.in

Received June 12, 2010; Accepted June 29, 2010; Published June 29, 2010


Copyright: © 2010 Selvaraj D, 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.
reveals that many O.sativa BZIP proteins have longer leucine zippers like A.thaliana and also shows similar dimerization property which is specified by attractive and repulsive g-e' interactions. Finally evolutionary relationship were analysed in plants by considering the BZIP protein.

### Materials and Methods

#### Data collection

The data set of *Indica Oryza sativa* of BZIP factors were obtained from the database of rice transcription factors (DART), http://drf.
The different plant BZIP factors were obtained from protein sequence database of National centre for Biotechnology Information (NCBI), http://www.ncbi.nlm.nih.gov/.

Pattern matching

The Pattern matching program Motifscan 3Dinsight is an integrated database and search tool for structure, function and sequence patterns of biomolecules which is used to identify the BZIP patterns existing between different plant sources. Two types of query of regular expressions were used in this database. The basic region expression [KR]-a-[RKSAQ]-N-[VL]-a-[SAQ]-x-[L]-[RKTAEN]-x-[R]-[S]-[RK] was found to be same in different plants. The numbers of motifs observed in the above plants were presented in Table 1.

Secondary structure predictions

Protein secondary structural elements were predicted using a new method called self optimized prediction method (SOPMA), which accurately predicts 69.5% of amino acid for the three state describing secondary structure (α-helix, β-sheet and coil). This tool works on the basis of neural network method (PHD) (Geourjon and Deleage, 1995).

Multiple alignment and phylogenetic tree analysis

Protein sequences were aligned with Clustal X Program and the number of heptad repeat distribution of BZIP in O.sativa was shown in the Figure 2. Allocation of Amino acid sequences of the Plants BZIP numbers were inferred using the Neighbor-Joining method. The optimal tree with the sum of branch length = 10.69915308 is shown. Phylogenetic analyses were conducted in MEGA4.

Results and Discussion

Amino acid content and leucine zipper length

The interaction of g-e position is characterized by charged amino acids like Arginine, serine, lysine, Proline and glycine. Pair of proline and glycine indicates the C- terminals. The plants like Antirrhinum majus, Capsicum annuum, Hyacinthus orientalis, Malus x domestica, Nicotiana tabacum, Phaseolus acutifolius, Phaseolus vulgaris, Triticum aestivum, Vitis vinifera shows triheptad repeats; Glycine max, Lycopersicon esculentum, Phaseolus tetragonolobus contains tetra repeats; Catharanthus roseus and Hyacinthus orientalis contains penta heptads; Spinacia oleracea shows hepta heptads these were shown in the Table 1. The kind of amino acids found in the a, d, e and g regions of O.sativa are found to be coiled coil arrangements which are known to regulate dimerization stability and specificity as shown in Figure 1. The number of leucine repeat distribution of BZIP in O.sativa was shown in the Figure 2. Allocation of Amino acid sequences of the Plants BZIP Domains were shown in Figure 4.

Biophysical analysis shows how the positions of amino acids contribute to dimerization specificity. From the present data, serine(S) in the second position of the heptad interacts with I, N, K and S at the ‘a’ position. These data indicates that serine contributes less to dimerization specificity than aliphatic amino acids, polar asparagines or charged lysine residues.

Phylogenetic analysis

The evolutionary relationships between the plants were evaluated by phylogenetic analysis of the aligned amino acids sequence of their BZIP domain. From the analysis BZIP factor 4 of Nicotiana tabacum is closely related with BZIP of Antirrhinum majus and Lycopersicon esculentum. The factor ATB2 BZIP of Glycine max is highly similar with BZIP-2 of Nicotiana tabacum and Capsicum annuum. BZIP factor 2 and 3 is found to be same in Phaseolus vulgaris. Factor 6 of Phaseolus vulgaris is closely related with taxon Triticum aestivum and Vitis vinifera putative ripening-related BZIP and Nicotiana tabacum. The plant Catharanthus roseus is not related with the above plants it has BZIP of G Box Binding protein, shown in the tree Figure 3. Evolutionary relationships of 23 taxa were inferred using the Neighbor-Joining method. The optimal tree with the sum of branch length = 10.69915308 is shown. Phylogenetic analyses were conducted in MEGA4.

Future Perspectives

Experimental work should be done through in vivo and in vitro methods which show different binding activities of bHLH and bZIP protein motifs. Yeast one Hybrid provides a satisfactory technique for in vivo testing of Protein – DNA interactions like bHLHZ targets with G-box. Through in vitro fluorescence anisotropy titrations protein homodimer are to be measured: E-box dissociation constants and circular dichorisim can be used to demonstrate the leucine zipper significance.

Conclusion

In this analysis dimerization partners of different plant BZIP proteins were analyzed by the neighbour – joining method (Saitou and Nei, 1987) using Molecular evolutionary Genetic Analysis tool (MEGA).

(Thomson et al., 1997). Phylogenetic relationship of different Plant BZIP proteins were analyzed by the neighbour – joining method (Saitou and Nei, 1987) using Molecular evolutionary Genetic Analysis tool (MEGA).
proteins were predicted and also observed for leucine zippers. The result reveals that plants *Phaseolus vulgaris*, *Capsicum annuum*, *Lycopersicon esculentum* and *Hyacinthus orientalis*, have three repeats where as *Spinaicia oleracea* has seven repeats of leucine. BZIP proteins were identified based on the presence of α helix breakers, proline, pair of glycines, presence of leucines in the d position, presence of charged amino acids in the g and e. Very few histidine residues are distributed in the plant source and suggest that such signaling system is absent. Phylogenetic analysis reveals that all BZIP proteins use the same amino acids to regulate dimerization specificity. Further experimental studies can be done to prove the dimerization property.

References


