Computational Biology and Chemistry 34 (2010) 268–283

Contents lists available at ScienceDirect

Computational Biology and Chemistry

journal homepage: www.elsevier.com/locate/compbiolchem

Research Article

In silico analysis of *cis*-acting regulatory elements in 5′ regulatory regions of sucrose transporter gene families in rice (*Oryza sativa* Japonica) and *Arabidopsis thaliana*

Omodele Ibraheem[a], Christiaan E.J. Botha[b], Graeme Bradley[a,*]

[a] *Plant Stress Response Group, Department of Biochemistry & Microbiology, University of Fort Hare, Private Bag X1314, Alice 5700, South Africa*
[b] *Plant Anatomy and Eco-physiology Research Laboratory, Department of Botany, Rhodes University, P.O. Box 94, Grahamstown 6140, South Africa*

ARTICLE INFO

Article history:
Received 26 May 2010
Received in revised form
14 September 2010
Accepted 27 September 2010

Key words:
Sucrose transporter
Gene regulation
Transcription factors
Phytohormones
Abiotic and biotic stress

ABSTRACT

The regulation of gene expression involves a multifarious regulatory system. Each gene contains a unique combination of *cis*-acting regulatory sequence elements in the 5′ regulatory region that determines its temporal and spatial expression. *Cis*-acting regulatory elements are essential transcriptional gene regulatory units; they control many biological processes and stress responses. Thus a full understanding of the transcriptional gene regulation system will depend on successful functional analyses of *cis*-acting elements.

Cis-acting regulatory elements present within the 5′ regulatory region of the sucrose transporter gene families in rice (*Oryza sativa* Japonica cultivar-group) and *Arabidopsis thaliana*, were identified using a bioinformatics approach.

The possible *cis*-acting regulatory elements were predicted by scanning 1.5 kbp of 5′ regulatory regions of the sucrose transporter genes translational start sites, using Plant CARE, PLACE and Genomatix Matinspector professional databases.

Several *cis*-acting regulatory elements that are associated with plant development, plant hormonal regulation and stress response were identified, and were present in varying frequencies within the 1.5 kbp of 5′ regulatory region, among which are; A-box, RY, CAT, Pyrimidine-box, Sucrose-box, ABRE, ARF, ERE, GARE, Me-JA, ARE, DRE, GA-motif, GATA, GT-1, MYC, MYB, W-box, and I-box. This result reveals the probable *cis*-acting regulatory elements that possibly are involved in the expression and regulation of sucrose transporter gene families in rice and *Arabidopsis thaliana* during cellular development or environmental stress conditions.

1. Introduction

The development of different cell types from an unchanging set of genes is ruled by physiological and biochemical processes that control gene activity. The selective expression of this set of genes directs plant development, cellular differentiation and responses to environmental stimuli (Franklin and Cande, 1999). Transcription in gene expression is the most essential step and has been extensively studied in cell and molecular biology (Wyeth and Albin, 2004). Gene specific regulation of transcription is of great significance for all phases of cellular functions.

Much of the regulatory portion of plant genes is located primarily in 1000 base pairs (bp) upstream of the transcriptional start site and this is generally referred to as the gene promoter region or 5′ regulatory region (Dean and Schmidt, 1995). The promoter region consists of specific DNA sequences and response elements that act in the recruitment of protein factors that facilitate transcription of the protein-coding region of the gene. These regulatory sequence elements located on the same strand as the coding region of the gene are called the *cis*-acting regulatory elements or the transcription factor binding sites (TFBs), and they determine the temporal and spatial expression of the gene. The *cis*-acting regulatory elements are specific short DNA sequence motifs of approximately 5–25 bp (Rani, 2007). The promoter regions are arbitrarily presented as a proximal promoter to −100 bp upstream of the TATA box and a distal promoter up to 2 kbp. The proximal sequence upstream of the gene contains the most important regulatory elements, while the distal sequence upstream may contain other regulatory elements, having weaker control on transcriptional gene regulation.

Transcriptional gene regulation is important for the function and development of all organisms. Living cells are products of intri-

* Corresponding author. Tel.: +27 40 602 2173; fax: +27 0866288637.
E-mail address: gbradley@ufh.ac.za (G. Bradley).

1476-9271/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compbiolchem.2010.09.003

cate gene expression programs that involve the transcription of countless genes. DNA binding transcription factors (TFs) are one of the important components of this network. Transcription is created by the interaction between transcription factors (TFs) that usually bind to the *cis*-acting regulatory elements in the gene with some additional co-factors to activate or repress gene transcription in response to change in the environment, as well as during development. A single TF seldomly controls gene expression, but precise combinations of many TFs are essential for the differential gene expression in higher organisms (Lloyd et al., 2001). This usually occurs on the promoters of each gene close to the transcription start site [(TSS); Qui, 2003]. Therefore, understanding of the *cis*-acting regulatory region bound by TFs that control gene expression will offer the essential information to build transcriptional regulatory systems (Qui, 2003).

The combination of computational methods with biology has advanced molecular modeling, and has also led to the development of large scale data inputs like microarrays, transcriptomes, proteomes and serial analysis gene expression (SAGE) (Qui, 2003; Jeffery et al., 2007). These high-throughput technologies have helped to accelerate the sequencing of varieties of genomes such as Arabidopsis, rice, human and rat in the past few years, and have also helped in the analysis of the expressional levels of many genes, consequently helping in gene discovery and pathway elucidation, contributing to the understanding of cell function.

Large significant portions of organismic DNA are encrypted to *cis*-acting regulatory programmes that control and coordinate gene expression, in which the transcriptional product depends on the cellular context and extracellular contributions. Many times external stimulus triggers a signal transduction pathway that leads to activation of several TFs that target enhancers/regulatory regions of genes, effecting their expression.

As well acknowledged, abiotic and biotic stress cause major losses in crop productivity worldwide. In plants, multiple signaling pathways regulate stress responses, and these often overlap with one another in response to different stresses. As an example, fungal infestation on plant gene expression profiles was found to overlap with those obtained for wounding (Durrant et al., 2000). Stress induced genes occur mostly at transcription levels, and their temporal and spatial expression regulatory patterns are vital part of the plant stress response system (Rushton and Somssich, 1998). Usually strongly related TFs have the possibility to activate or suppress the transcription of genes by interacting with the *cis*-acting regulatory sequences that respond to a particular stress. Plants have been attested to devote a large portion of their genome capacity to transcription, with the Arabidopsis genome coding in excess of 1500 transcription factors (Riechmann et al., 2000).

Plant hormones such as ethylene, salicylic acid (SA), jasmonic acid (JA) and abscisic acid (ABA) are important regulators of stress responsive pathways. Ethylene and JA are commonly associated with defence pathways activated by pathogen infection and wounding (Zimmerli et al., 2004), while ABA is well known to regulate response to abiotic factors such as salinity and drought (Zhu, 2002). Recently, ABA has been implicated in the establishment of compatible interactions between fungal pathogens and host plants (Mohr and Cahill, 2003). Exogenous applications of ethylene (Jacobs et al., 1999), JA (Zimmerli et al., 2004) and ABA (Finkelstein and Gibson, 2002) have been reported to elicit aspects of the host response to stress stimuli.

Many stress-associated *cis*-acting regulatory elements that activate transcription in response to salinity, drought, wounding and pathogen infection have been identified in plants (Higo et al., 1999; Singh et al., 2002; Rani, 2007). Web-based databases such as: Plant CARE, PLACE and Genomatix Matinspector professional, have provided a convenient and easy way to search for these previously identified promoter motifs in DNA sequences. They are databases of plant *cis*-acting regulatory elements and they provide access tools for in silico analysis of regulatory sequences. They are a collection of short DNA motif sequences present in plant *cis*-acting regulatory regions, obtained from earlier published research works, and also from article reviews on the regulatory regions of various plant genes. The variations that occur in these motifs in other genes or plants in subsequent research reports are also included (Higo et al., 1999; Lescot et al., 2002; Rani, 2007).

Many sucrose transporters have been cloned from plant tissues (Sauer, 2007), but little is known about the expression and the activities of these transporters to environmental stress. To date, no work has been published on the transcription factors/*cis*-acting regulatory elements that regulate the expression of sucrose transporter genes during stress. Therefore the focus of this study was to use prediction of available bioinformatic tools to reveal a comprehensive description of *cis*-acting regulatory elements that are present within the 5′ regulatory region of the DNA sequences of sucrose transporter gene families in rice (*Oryza sativa* Japonica cultivar-group) and *Arabidopsis thaliana*; as a source of potentially useful information for the prediction of expression and regulation of these genes. This analysis is the first step towards uncovering transcription regulatory interactions of these genes during development or under environmental stress conditions.

2. Method

2.1. Identification of the 5′ regulatory region of sucrose transporter genes

The complete nucleotide sequence and the coding domain sequence (CDS) of the sucrose transporter gene family in rice as registered in GenBank by Aoki et al. (2003), under the accession number: OsSUT1 (AF280050), OsSUT2 (AB091672), OsSUT3 (AB071809), OsSUT4 (AB091673), OsSUT5 (AB091674) were obtained from National Center for Biotechnology Information (NCBI) internet database site (http://www.ncbi.nih.nim.gov). Similarly, the nucleotide sequence and the coding domain sequence (CDS) of the sucrose transporter gene family in *Arabidopsis thaliana* as registered in Genbank under the accession number: AtSUC1 (At1g71880), AtSUC2 (At1g22710), AtSUC3/AtSUT2 (At2g02860), AtSUC4 (At1g09960), AtSUC5 (At1g71890), AtSUC6 (At5g43610), AtSUC7 (At1g66570), AtSUC8 (At2g14670), AtSUC9 (At5g06170); (Sauer, 2007) were also obtained from National Center for Biotechnology Information (NCBI) internet database site.

Individual sequences of rice and *Arabidopsis thaliana* sucrose transporter gene families were confirmed against the *Oryza sativa* (Japonica cultivar-group) genome and *Arabidopsis thaliana* genome respectively, using the Basic Local Alignment Tool (BLASTN; http://blast.ncbi.nlm.nih.gov/Blast.cgi). Locus link was used to identify a genomic sequence of 1.5 kbp extending 5′ from the translation start site of each sucrose transporter gene to the next gene. These sequences were used for the computational analysis.

2.2. Cis-acting regulatory elements analysis

Using database associated search tools, 1.5 kbp of 5′ regulatory region of each sucrose transporter gene in rice and *Arabidopsis thaliana* were scanned for the presence of putative *cis*-acting regulatory elements identical with or similar to the motifs registered in Plant CARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/cgi-bin/CallMat_IE55.htpl), PLACE (http://www.dna.affrc.go.jp/PLACE/signalscan.html) and Genomatix Matinspector professional (http://www.genomatix.de/cgi-bin/matinspector_prof/mat_fam.p1).