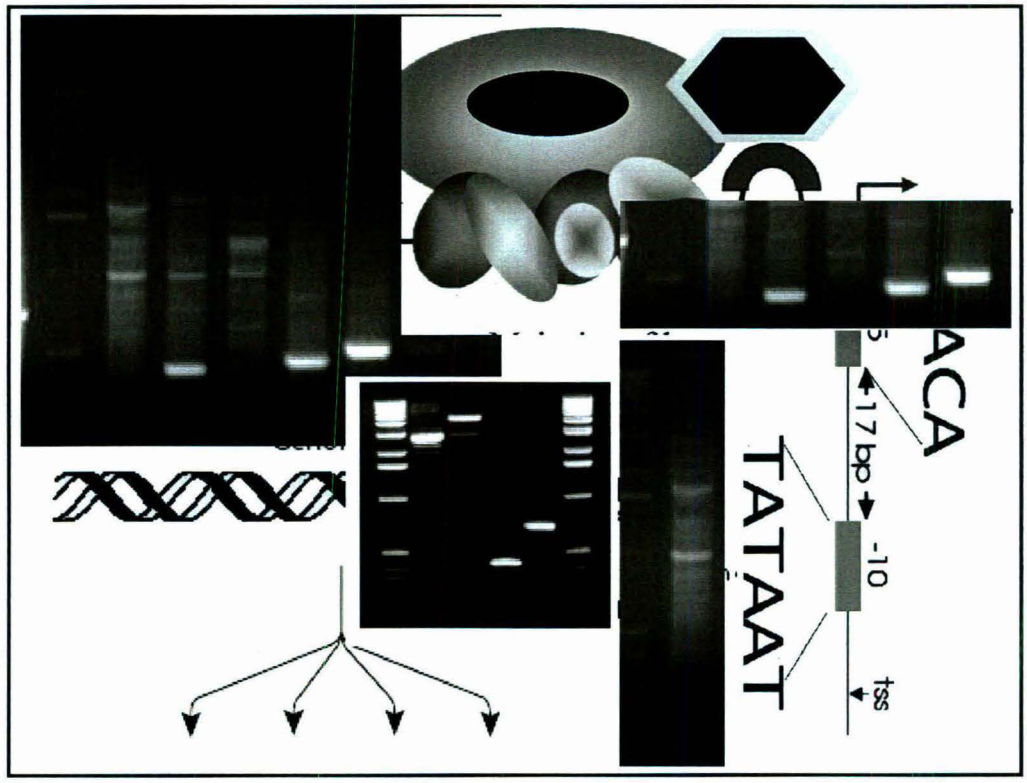
Chapter 4



Promoter Cloning and in silico Analysis

CHAPTER 4

Promoter cloning and *in-silico* analysis

4.1 Introduction

The promoter contains specific DNA sequences that are recognized by proteins known as transcription factors. These factors bind to the promoter sequences, recruiting RNA polymerase, the enzyme that synthesizes the RNA from the coding region of the gene. Promoters represent critical elements that can work in concert with other regulatory regions (enhancers, silencers, boundary elements/insulators) to direct the level of transcription of a given gene. Eukaryotic promoters are extremely diverse and are difficult to characterize. They typically lie upstream of the gene and can have regulatory elements several kilobases away from the transcriptional start site. Because in eukaryotes, the transcriptional complex can cause the DNA to bend back on itself, which allows for placement of regulatory sequences far from the actual site of transcription.

Understanding how the regulation of gene networks is orchestrated is an important challenge for characterizing complex biological processes. Gene transcription is regulated in part by nuclear factors that recognize short DNA sequence motifs, called transcription factor binding sites, in most cases located upstream of the gene coding sequence in promoter and enhancer regions. Genes expressed in the same tissue under similar conditions often share a common organization of at least some of these regulatory binding elements. In this way the organization of promoter motifs represents a "footprint" of the transcriptional regulatory mechanisms at work in a specific biological context and thus provides information about signal and tissue specific control of expression. Therefore analysis of promoters provides a crucial link between the static nucleotide sequence of the genome and the dynamic aspects of gene regulation and expression.

4.2 Methods

4.2.1 Promoter isolation by Genome Walking

Genome walking is a simple method for finding unknown genomic DNA sequences adjacent to a known sequence such as a cDNA. The promoter of *SlAMS* was isolated using Universal Genomewalker™ Kit (Clontech, USA). From this kit a pool of uncloned, adaptor-ligated genomic DNA fragments were obtained which are referred to as libraries. These libraries are then used for isolation of gene specific promoter.

1. First step is isolation of high quality genomic DNA better than suitable for southern blot and conventional PCR. Therefore good quality DNA was isolated from tomato using the protocol described in the section 3.3.5.6.1. The quality of genomic DNA

was determined by resolving it on agarose/EtBr gel. The DNA obtained was intact as no smear was observed.

2. In four different 1.5 ml sterile tubes, four digestion reactions were set up using the enzymes *Dra* I, *Eco*R V, *Pvu* II and *Stu* I. For our convenience we choose five enzymes for digestion. All these enzymes produce blunt ends. In each reaction following components were combined.

Genomic DNA	25 µl
Restriction enzyme	8 µl
Restriction enzyme buffer	10 µl
Deionized water	57 µl
Total volume	100 µl

The contents were mixed gently and incubated at 37^0C for 2 hr. The tubes were tapped gently and again kept for 16-18 hr. From each tube 5 µl reaction mix was checked for digestion on 0.5% agarose/EtBr gel.

3. To each of the reaction tube, an equal amount (95 µl) of phenol was added and slowly vortexed for 10 sec. After a brief spin aqueous layer was transferred to a new tube and again the above step was repeated to remove protein contamination.
4. After the second extraction, 2 volumes (190 µl) of ice cold 95% ethanol, 1/10 volumes (9.5 µl) of 3M NaOAc, and 20 µg of glycogen was added and vortexed slowly for 10 sec.
5. To pellet the digested DNA, the tubes were centrifuged at 15,000 rpm for 10 min and the supernatant was decanted. The pellets obtained were washed with in 100 µl of ice cold 80% ethanol and centrifuged at 15,000 rpm for 5 min.
6. The supernatant was decanted, pellet was air dried and dissolved in 20 µl of TE pH 7.5. After a slow speed vortex for 5 sec, 1 µl of the Digested DNA quality & quantity was checked on a 0.5% agarose/ EtBr gel.
7. For ligation, 4 µl of each digested and purified DNA was taken in 0.5 ml tubes and to each of the four tubes the following components were added 1.9 µl Genomewalker adaptor (25µM), 1.6 µl 10X Ligation buffer and 0.5µl T4 DNA ligase (6 units/µl). This reaction was incubated overnight at 16^0C.
8. Next day, to stop the reaction, the tubes were incubated at 70^0C for 5 min and the ligated DNA was diluted with 72 µl of TE (pH 7.4).
9. The promoter regions of the desired gene were amplified from the five adaptor ligated DNA library. PCR was done separately for each library with AP1 primer, and gene specific primers

10. Some of the important points during design of gene specific primers are, they should be 26-30 mer in length and GC content should be 40-60%. The nested gene specific primer should be internal or upstream to the GSP1. There should not be more than three G's and C's in the last six positions at the 3' end of the primer.

    The primary PCR reaction mix had the following components

Deionized water	40 µl
10 X Advantage 2 PCR buffer	5µl
dNTP mix (10µM each)	1µl
AP1 (10 µM)	1µl
DNA library	1µl
GSP1(10 µM)	1µl
Advantage 2 ploymerase mix	1µl

The cycling conditions were

7 cycles
94°C 25 sec
72°C 3 min
32 cycles
94 °C 25 sec
67 °C 3 min
67°C for additional 7 min and the primary PCR product was run on 1.5% Agarose/EtBr gel.

11. After the primary PCR, 1 µl of it was diluted to 50 µl to perform nested amplification using AP2 and GSP2 primer. The secondary PCR reaction mixture had the same components as that of the primary except for AP1 which was replaced by AP2 and the template was, diluted primary PCR product.
12. The amplified product was eluted after running on 1.5% agarose/EtBr gel and further cloned in to pGEM-T Easy vector. The clones were sequenced and analyzed *in-silico* to identify the putative regulatory upstream elements.

4.3 Results

GenomeWalker kit was used to isolate upstream elements of the cDNA (Fig. 4.1). Five genomic libraries were prepared by digesting with *Pvu* II, *Xmn* I *Msc* I, *Dra* I and *Ssp* I (Fig 4.2A) and ligating the adapters after purifying the digested DNA. Two gene specific primers were designed using the criteria mentioned above. The primary amplification was done using AP1 and GSP1 (5' TGA CTT CCA CTA CCC ACA ACC CAC 3') (Fig. 4.2B). The secondary PCR was done by diluting 1µl of the primary to 50 µl and using AP2 and GSP2 (5' CCA CCA TTG CTG AGA GTG TTC ACT GAA 3') (Fig. 4.2 C). The fragment obtained was gel eluted and cloned in T-vector. After sequencing it was aligned with the cDNA to identify the overlapping region. The promoter region which we amplified was ~ 1.2 Kb in length and was amplified from *Mse* I library (Fig. 4.3). Further the sequence was analyzed by two servers (PLACE, Plant Cis acting regulatory DNA elements) (Table 4.1) and PlantCARE

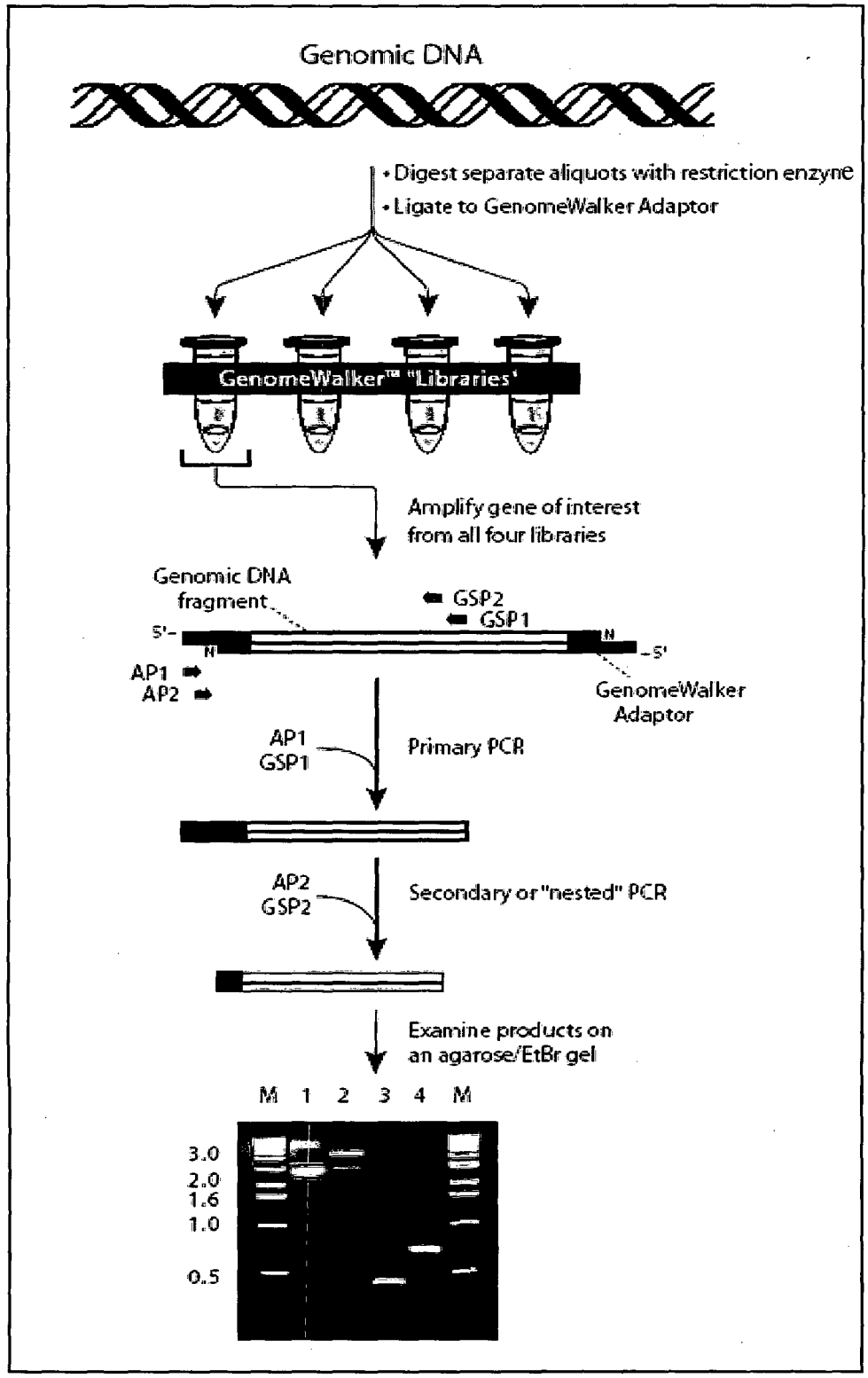


Figure 4.1 Flow chart of the Genome Walker protocol.

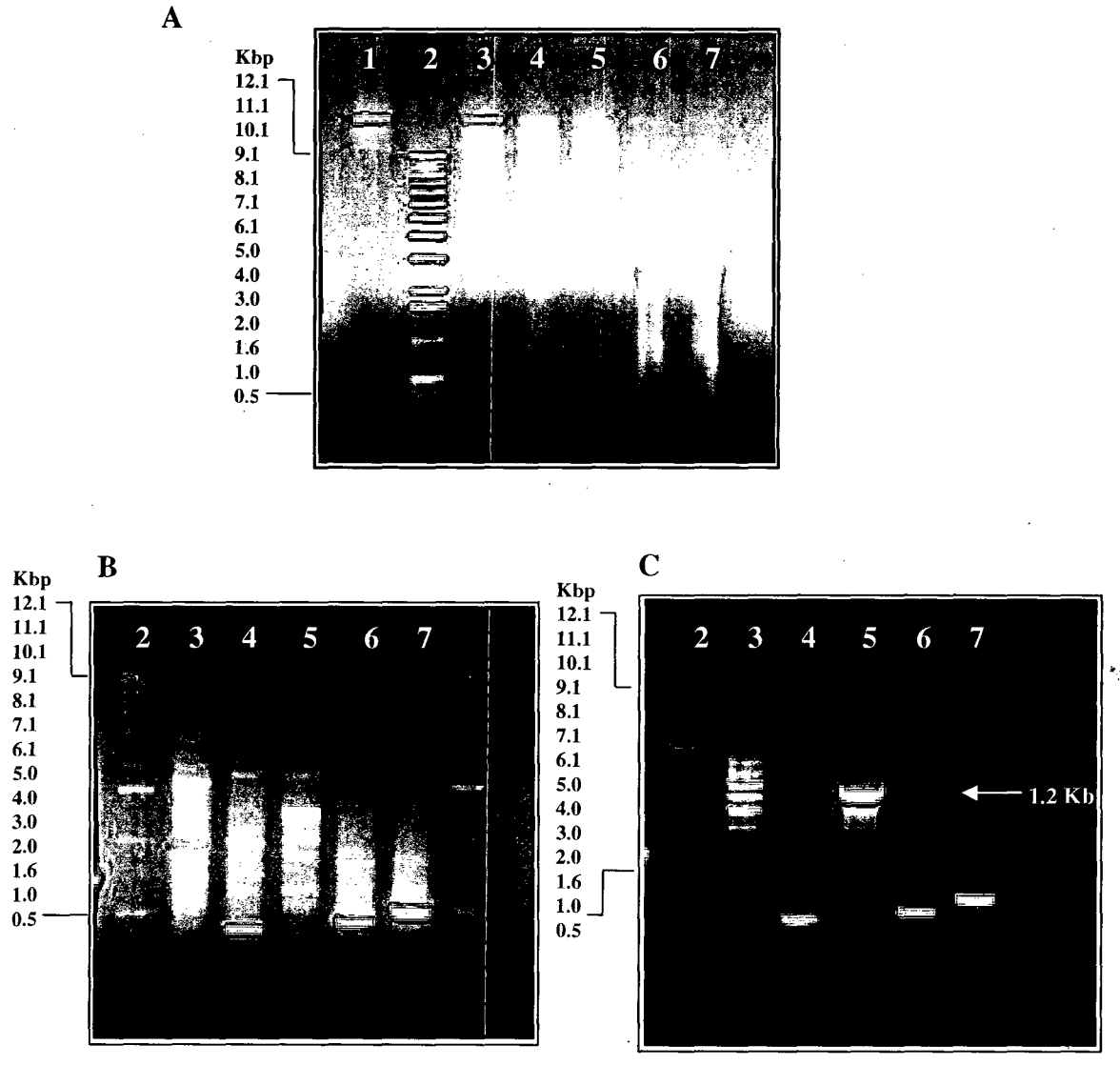


Figure 4.2 Promoter isolation by PCR based Genome walking. (A) Digestion of genomic DNA with *Pvu* II (lane 3), *Xmn* I (lane4), *Msc* I (lane 5), *Dra* I (lane 6) and *Ssp* I (lane 7). Adapters were ligated to the digested fragments and referred as libraries (Fig 4.1). (B) Primary PCR with AP1 and GSP1 using the libraries constructed as mentioned in A. (C) Secondary PCR with the AP2 and GSP2 primers using diluted primary PCR product as the template. The arrow indicates the 1.2 kb fragment amplified from *Mse* I library, which was cloned in pGEM-T easy vector and sequenced. Lane 1 is undigested genomic DNA of tomato and lane 2 is DNA marker (Invitrogen, 1Kb ladder).

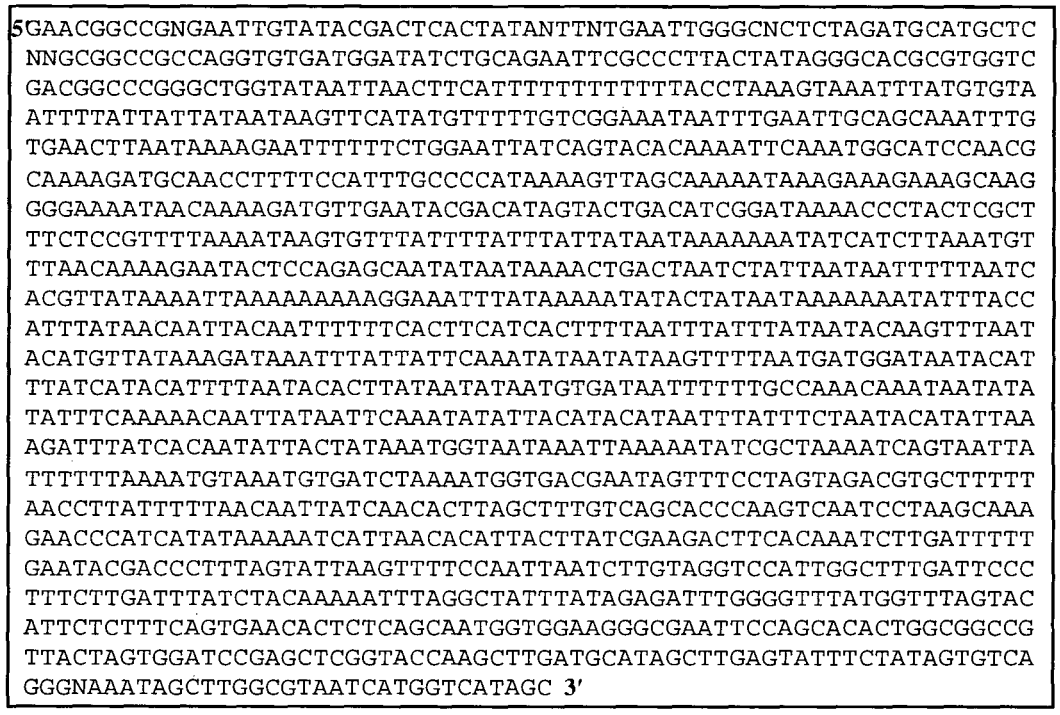


Figure 4.3 Promoter sequence of the *SlAMS* obtained by genome walking. Sequence in black represents the promoter region and sequence in blue represents 5' UTR of the gene.

(Plant cis-acting regulatory elements) available for plants. The PLACE is a database of motifs found in plant cis-acting regulatory DNA elements, all from previously published reports. In addition to the motifs originally reported, their variations in other genes or in other plant species reported later are also compiled. The PLACE database (Higo et al., 1999) also contains a brief description of each motif and relevant literature with PubMed ID numbers. The web form can be accessed at http://www.dna.affrc.go.jp/PLACE/. With in the 1233 bp upstream regulatory region of *SlAMS,* obtained by PCR-based chromosome walking. The sequence analysis of the promoter region predicted some putative cis-elements related to abiotic and abiotic stress response, many light response elements, CGCG box involved in multiple signal transduction pathway, ABRE elements, GA response elements, GCC elements present in PR genes, auxin response elements and many more are described in the table 4.1.

Promoter analysis was also done using Plant CARE, which identify the elements present in the sequence by comparing with the already known plant elements. PlantCare is a database of plant cis-acting regulatory elements, enhancers and repressors. Data about the transcription sites are extracted mainly from the literature, supplemented with an increasing number of *in-silico* predicted data. Apart from a general description for specific transcription factor sites, levels of confidence for the experimental evidence, functional information and the position on the promoter are given as well. Using PlantCARE (Lescot et al., 2002) we could find motifs like Auxin responsive element found in parC gene in tobacco (Sakai et al., 1996). Box 4 motif, which is a cis-acting element present in PAL gene of parsely (Lois et al., 1989), light responsive elements like G-Box, GT1 motif, MNF1 elements were also identified. A sequence similar to protein binding site (called BoxIII) present in rbcs-3A promoter region in pea (Kuhlemeier et al., 1987) was another light responsive element identified. Motif similar to cis-acting regulatory element involved in MeJA-response (Methyl Jasmonate) similar to the one present in lipoxygenase isoenzyme in *Hordium vulgare* (Rouster et al., 1997) was also identified in the promoter. Ethylene-responsive element, similar to the one present in GST gene of carnation, was identified by the program (Itzhaki and Woodson, 1993). Wound responsive element similar to *Brassica oleracea* S gene family receptor -like kinase which is specific to wounding was identified (Pastuglia et al., 1997) in the *SlAMS* upstream elements. Regulatory elements similar to one involved in circadian control like tomato chlorophyll a/b-binding protein and Cab-1A gene, (Pichersky et al., 1985) were present. Another interesting regulatory element similar to 5' UTR Py-rich stretch which is know to confer high transcription levels without the need for other upstream cis elements except for a TATA-box was identified (Daraselia et al., 1996). The web form of the plant CARE can be accessed from http://bioinformatics.psb.ugent.be/webtools/plantcare.

Factor or Site Name	Loc.(Str.)	Signal Sequence	ID	Function and Reference
-10PEHVPSBD	513 (-)	TATTCT	S000392	Involved in the expression of the plastid gene psbD which encodes a photosystem II reaction center chlorophyll-binding protein that is activated by blue, white or UV-A light. Thum et al., 2001
ABRELATERD1	998 (+) 567 (-)	ACGTG	S000414	ABRE-like sequence (from -199 to -195) required for etiolation-induced expression of erd1 (early responsive to dehydration) in Arabidopsis. Simpson et al., 2003
ABRERATCAL	116 (+) 117 (-)	MACGYGB	S000507	ABRE-related sequence" or "Repeated sequence motifs" identified in the upstream regions of Ca(2+)-responsive upregulated genes; see also ABRE; M=C/A; Y=T/C; B=T/C/G; Kaplan et al., 2006
ACGTATERD1	568 (+) 998 (+) 998 (-)	ACGT	S000415	ACGT sequence (from -155 to -152) required for etiolation-induced expression of erd1 (early responsive to dehydration) in Arabidopsis; Simpson et al., 2003
AGL2ATCONSENSUS	897 (-) 898 (+) 898 (-)	NNWNCCAWW WWTRGWWAN	S000339	Binding consensus sequence of Arabidopsis (A.t.) AGL2 (AGAMOUS-like 2); AGL2 contains MADS domain; AGL2 binds DNA as a dimmer. Huang et al., 1996.
AMYBOX1	386 (+) 506 (+)	TAACARA	S000020	"amylase box"; Conserved sequence found in 5'-upstream region of alpha-amylase gene of rice, wheat, barley; Huang et al., 1990
AMYBOX2	82 (-) 744 (-)	TATCCAT	S000021	"amylase box"; "amylase element"; Conserved sequence found in 5'upstream region of alpha-amylase gene of rice, wheat, barley; Hwang et al., 1998
ARR1AT	883 (+) 1127 (+) 546 (-)	NGATT	S000454	"ARR1-binding element" found in Arabidopsis; ARR1 is a response DE regulator; N=G/A/C/T; AGATT is found in the promoter of rice DE non-symbiotic haemoglobin-2 (NSHB) gene (Ross et al., 2004); Ross et al., 2004
ASF1MOTIFCAMV	977 (+)	TGACG	S000024	"ASF-1 binding site" in CaMV 35S promoter; ASF-1 binds to two TGACG motifs; Found in HBP-1 binding site of wheat histone H3 gene; TGACG motifs are found in many promoters and are involved in transcriptional activation of several genes by auxin and/or salicylic acid; May be relevant to light DE regulation; Binding site of tobacco TGA1a; TGA1a and b show homology to CREB; TGA6 is a new member of the TGA family; Abiotic and biotic stress differentially stimulate "as-1 element" activity; Redman et al., 2002

BIHD1OS	1043 (+) 416 (-)	TGTCA	S000498	Binding site of OsBIHD1, a rice BELL homeodomain transcription factor; Luo et al., 2005
CAATBOX1	527 (+) 639 (+) 645 (+)	CAAT	S000028	"CAAT promoter consensus sequence" found in legA gene of pea; Shirsat et al., 1989
CACTFTPPCA1	26 (+) 1102 (+) 174 (-)	YACT	S000449	Tetranucleotide (CACT) is a key component of Mem1 (mesophyll expression module 1) found in the cis-regulatory element in the distal region of the phosphoenolpyruvate carboxylase (ppcA1) of the C4 dicot F. trinervia Gowik et al., 2004
CARGCW8GAT	875 (+) 875 (-)	CWWWWWWW WG	S000431	A variant of CArG motif (see S000404), with a longer A/T-rich core; Binding site for AGL15 (AGAMOUS-like 15) Folter and Angenent 2006
CATATGGMSAUR	211 (+) 211 (-)	CATATG	S000370	Sequence found in NDE element in soybean (G.m.) SAUR (Small Auxin-Up RNA) 15A gene promoter; Involved in auxin responsiveness. Xu et al., 1997
CBFHV	124 (+) 124 (-)	RYCGAC	S000497	Binding site of barley (H.v.) CBF1, and also of barley CBF2; CBF = C-repeat (CRT) binding factors; CBFs are also known as dehydration-responsive element (DRE) binding proteins (DREBs); Svensson et al., 2006
CCAATBOX1	1162 (+) 40 (-) 1182 (-)	CCAAT	S000030	Common sequence found in the 5'-non-coding regions of eukaryotic genes; "CCAAT box" found in the promoter of heat shock protein DE genes; Located immediately upstream from the most distal HSE of the promoter; "CCAAT box" act cooperatively with HSEs to increase the hs promoter activity; Wenkel et al., 2006
CGACGOSAMY3	126 (+)	CGACG	S000205	"CGACG element" found in the GC-rich regions of the rice (O.s.) Amy3D and Amy3E amylase genes, but not in Amy3E gene; May function as a coupling element for the G box element; Hwang et al., 1998
CGCGBOXAT	117 (+) 117 (-)	VCGCGB	S000501	"CGCG box" recognized by AtSR1-6 (Arabidopsis thaliana signal-responsive genes); Multiple CGCG elements are found in promoters of many genes; Ca++/calmodulin binds to all AtSRs; Yang and Poovaiah 2002
CIACADIANLELHC	1053 (+)	CAANNNNATC	S000252	Region necessary for circadian expression of tomato (L.e.) Lhc DE gene; Piechulla et al., 1998

CPBCSPOR	869 (-)	TATTAG	S000491	The sequence critical for Cytokinin-enhanced Protein Binding in vitro, found in -490 to -340 of the promoter of the cucumber (CS) POR (NADPH-protochlorophyllide reductase) gene. Fusada et al., 2005
CRTDREHVCBF2	124 (+) 124 (-)	GTCGAC	S000411	Preferred sequence for AP2 transcriptional activator HvCBF2 of DE barley; "Core CRT/DRE motif"; HvCBF2 bound to a (G/a)(T/c)CGAC core motif (Xue, 2003); DNA binding is regulated by temperature; This motif was erroneously ID-labeled in PLACE db as CDTDREHVCBF2. Xue, 2003
CURECORECR	288 (+) 412 (+) 288 (-)	GTAC	S000493	GTAC is the core of a CuRE (copper-response element) found in Cyc6 and Cpx1 genes in Chlamydomonas; Also involved in oxygen-response of these genes. Kropat et al., 2005
DOFCOREZM	172 (+) 264 (+) 318 (+)	AAAG	S000265	Core site required for binding of Dof proteins in maize (Z.m.); Dof proteins are DNA binding proteins, with presumably only one zinc finger, and are unique to plants; Four cDNAs encoding Dof proteins, Dof1, Dof2, Dof3 and PBF, have been isolated from maize; PBF is an endosperm specific Dof protein that binds to prolamin box; Maize Dof1 enhances transcription from the promoters of both cytosolic orthophosphate kinase (CyPPDK) and a DE non-photosynthetic PEPC gene; Maize Dof2 supressed the C4PEPC promoter; Yanagisawa, 2000
DPBFCOREDCDC3	74 (-)	ACACNNG	S000292	A novel class of bZIP transcription factors, DPBF-1 and 2 (Dc3 promoter-binding factor-1 and 2) binding core sequence; Found in the carrot (D.c.) Dc3 gene promoter; Dc3 expression is normally embryo-specific, and also can be induced by ABA; The Arabidopsis abscisic acid response gene ABI5 encodes a bZIP transcription factor; abi5 mutant have a pleiotropic defects in ABA response; ABI5 regulates a subset of late embryogenesis-abundant genes; GIA1 (growth-insensitivity to ABA) is identical to ABI5. Finkelstein and Lynch, 2000
E2FCONSENSUS	97 (+) 337 (+)	WTTSSCSS	S000476	"E2F consensus sequence" of all different E2F-DP-binding motifs that were experimentally verified in plants. Vandepoele et al., 2005
EBOXBNNAPA	74 (+) 211 (+) 300 (+)	CANNTG	S000144	E-box of napA storage-protein gene of Brassica napus (B.n.); (Myc consensus: CANNTG); This sequence is also known as RRE (R response element) Hartmann et al., 2005

EECCRCAH1	236 (+)	GANTTNC	S000494	"EEC"; Consensus motif of the two enhancer elements, EE-1 and EE-2, both found in the promoter region of the Chlamydomonas Cah1 (encoding a periplasmic carbonic anhydrase); Binding site of Myb transcription factor LCR1 Yoshioka et al., 2004
ERELEE4	296 (+) 821 (+) 838 (+)	AWTTCAAA	S000037	"ERE (ethylene responsive element)" of tomato (L.e.) E4 and carnation GST1 genes; GST1 is related to senescence; Found in the 5'-LTR region of TLC1.1 retrotransposon family in Lycopersicon chilense (Tapia et al.); ERE motifs mediate ethylene-induced activation of the U3 promoter region; Rawat et al., 2005
GAREAT	386 (+) 506 (+)	TAACAAR	S000439	GARE (GA-responsive element); Occurrence of GARE in GA-inducible, GA-responsible, and GA-nonresponsive genes found in Arabidopsis seed germination was 20, 18, and 12%, respectively Ogawa et al., 2003
GATABOX	85 (+) 424 (+) 706 (+)	GATA	S000039	"GATA box"; GATA motif in CaMV 35S promoter; Binding with ASF-2 Three GATA box repeats were found in the promoter of Petunia (P.h.) chlorophyll a/b binding protein, Cab22 gene; Required for high level, light regulated, and tissue specific expression; DE Conserved in the promoter of all LHCII type I Cab genes; Benfey et al., 1990
GCCCORE	69 (+)	GCCGCC	S000430	Core of GCC-box found in many pathogen-responsive genes such as PDF1.2, Thi2.1, and PR4; Has been shown to function as ethylene-responsive element; Appears to play important roles in regulating jasmonate-responsive gene expression; Tomato Pti4 (ERF) regulates defence-related gene expression via GCC box and non-GCC box cis elements (Myb1 (GTTAGTT) and G-box(CACGTG)); Chakravarthy et al., 2003
GT1CONSENSUS	225 (+) 380 (+) 381 (+)	GRWAAW	S000198	Consensus GT-1 binding site in many light-regulated genes, e.g., RBCS from many species, PHYA from oat and rice, spinach RCA and PETA, and bean CHS15; R=A/G; W=A/T; For a compilation of related GT elements and factors, see Villain et al. (1996); GT-1 can stabilize the TFIIA-TBP-DNA (TATA box) complex; The activation mechanism of GT-1 may be achieved through direct interaction between TFIIA and GT-1; Binding of GT-1-like factors to the PR-1a promoter influences the level of SA-inducible gene expression; Zhou, 1999

GT1GMSCAM4	271 (-) 649 (-)	GAAAAA	S000453	"GT-1 motif" found in the promoter of soybean (Glycine max) CaM isoform, SCaM-4; Plays a role in pathogen- and salt-induced SCaM-4 gene expression. Park et al., 2004
GTGANTG10	79 (+) 252 (+) 790 (+)	GTGA	S000378	"GTGA motif" found in the promoter of the tobacco (N.t.) late pollen gene g10 which shows homology to pectate lyase and is the putative homologue of the tomato gene lat56; Located between -96 and -93. Rogers et al., 2001
IBOX	1104 (-)	GATAAG	S000124	"I box"; "I-box"; Conserved sequence upstream of light-regulated genes; Sequence found in the promoter region of rbcS of tomato and Arabidopsis; I box (Giuliano et al. 1988); Binding site of LeMYB1, that is a member of a novel class of myb-like proteins; LeMYB1 act as a transcriptional activator. Rose et al., 1999
IBOXCORE	424 (+) 706 (+) 747 (+)	GATAA	S000199	"I box"; "I-box"; Conserved sequence upstream of light-regulated genes; Conserved sequence upstream of light-regulated genes of both monocots and dicots. Terzaghi and Cashmore, 1995
L1BOXATPDF1	752 (-)	TAAATGYA	S000386	"L1 box" found in promoter of Arabidopsis thaliana (A.t.) PROTODERMAL FACTOR1 (PDF1) gene; Located between -134 and -127; Involved in L1 layer-specific expression; L1-specific homeodomain DE protein ATML can bind to the "L1 box"; Y=C/T; A cotton fiber gene, RD22-like 1 (RDL1), contains a homeodomain binding L1 box and a MYB binding motif (Wang et al., 2004); HDZip IV. Ohashi et al., (2003)
LTRECOREATCOR15	222 (-)	CCGAC	S000153	Core of low temperature responsive element (LTRE) of cor15a gene in Arabidopsis (A.t.); A portion of repeat-C (C-repeat), TGGCCGAC, which is repeated twice in cor15a promoter (Baker et al., 1994); ABA responsiveness; Involved in cold induction of BN115 gene from winter Brassica napus; LTRE; Light signaling mediated by phytochrome is necessary for cold- or drought- induced gene expression through the C/DRE in Arabidopsis; Kim et al., (2002)
MARABOX1	468 (-)	AATAAAYAAA	S000063	"A-box" found in SAR(scaffold attachment region; or matrix attachment region, MAR). Gasser et al., 1989

MARTBOX	156 (+) 157 (+) 158 (+)	TTWTWTTWTT	S000067	"T-Box"; Motif found in SAR (scaffold attachment region; or matrix attachment region, MAR). Gasser et al., 1989
MYBCOREATCYCB1	2 (+)	AACGG	S000502	"Myb core" in the 18 bp sequence which is able to activate reporter gene without leading to M-phase-specific expression, found in the promoter of Arabidopsis thaliana cyclin B1:1 gene; the 18 bp sequence share homology with a sequence found in the N. sylvestris cyclin B1 promoter Planchais et al., 2002
MYBGAHV	386 (+)	TAACAAA	S000181	Central element of gibberellin (GA) response complex (GARC) in high-pI alpha-amylase gene in barley (H.v.); Similar to c-myb and v-myb consensus binding site; GAmyb binds specifically to the TAACAAA box in vitro; GAmyb is the sole GA-regulated transcriptional factor required for transcriptional activation of the high-pI alpha-amylase; GARC consist of the pyrimidine, TAACAAA and TATCCAC boxes; GARE in RAmy1A gene; GARE and pyrimidine box in RAmy1A are partially involved in sugar repression. Gubler et al., 1999
PRECONSCRHSP70A	203 (-)	SCGAYNRNNNN NNNNNNNNNN NHD	S000506	Consensus sequence of PRE (plastid response element) in the promoters of HSP70A in Chlamydomonas; Involved in induction of HSP70A gene by both MgProto and light
RAV1BAT	74 (-)	CACCTG	S000315	Binding consensus sequence of an Arabidopsis (A.t.) transcription DE factor, RAV1; RAV1 specifically binds to DNA with bipartite DE sequence motifs of RAV1-A (CAACA) and RAV1-B (CACCTG); RAV1 DE protein contain AP2-like and B3-like domains; The AP2-like and DE B3-like domains recognize the CAACA and CACCTG motifs, DE respectively; The expression level of RAV1 were relatively high DE in rosette leaves and roots Kagaya et al., 1999
REBETALGLHCB21	422 (+)	CGGATA	S000363	"REbeta" found in Lemna gibba Lhcb21 gene promoter; Located at -114 to -109; A GATA sequence created at a position six nucleotides upstream could replace the function of REbeta; Required for phytochrome regulation. Degenhardt and Tobin, 1996
RYREPEATGMGY2	54 (-)	CATGCAT	S000105	"RY repeat motif (CATGCAT)"; Present in the 5' region of the DE soybean (G.m.) glycinin gene (Gy2). Lelievre et al., 1992

TATCCAYMOTIFOSRAMY3D	82 (-) 744 (-)	TATCCAY	S000256	"TATCCAY motif" found in rice (O.s.) RAmy3D alpha-amylase gene promoter; Y=T/C; a GATA motif as its antisense sequence; TATCCAY motif and G motif are responsible for sugar repression Rubio-Somoza et al ., 2006
UP2ATMSD	428 (+)	AAACCCTA	S000472	"Up2" motif found in 193 of the 1184 up-regulated genes after DE main stem decapitation in Arabidopsis; Tatematsu et al., 2005
WBOXNTCHN48	540 (+)	CTGACY	S000508	"W box" identified in the region between -125 and -69 of a tobacco class I basic chitinase gene CHN48; NtWRKY1, NtWRKY2 and NtWRKY4 bound to W box; NtWRKYs possibly involved in elicitor-respsonsive transcription of defense genes in tobacco; Y=C/T; see also S000442 (TGACT) and S000447 (TGAC); Yamamoto et al., 2004

Table 4.1 The putative regulatory cis-elements identified in *SlAMS* promoter by PLACE (Plant Cis acting regulatory DNA elements)

4.4 Discussion

The isolation and characterization of fruit-specific promoters are critical for the manipulation of the nutritional value and quality of fruits by genetic engineering. The analysis of regulatory sequences of many ripening related gene promoters is a challenging task. As the promoter region of these genes comprise highly divergent sequences visualized in modular terms whose activities are controlled by the combinatorial association of multiple proteins. Knowledge about this complex interaction that controls gene expression at the transcriptional level is still very limited. This makes it very difficult to predict the activity of a promoter in a different physiological context, either in different organ or species (Agius et al., 2005).

The *SlAMS* gene showed differential expression profiles during different stages of fruit ripening. The promoter region of the *SlAMS* gene contained several putative functional cis-elements, which may be involved in perceiving stimulus from different plant hormones and environmental stresses. In order to obtain more information on the regulatory mechanism of expression of *SlAMS,* promoter region was isolated by genome walking. In addition to the ubiquitous elements including TATA CAAT boxes, the *SlAMS* promoter region contains sequences similar to the regulatory cis-elements found in other plant genes. A potential abscisic acid responsive element (ABRE) was found which was required for expression of erd1 (early response to dehydration) in *Arabidopsis.* An ethylene-responsive element (ERE) was located in the promoter region giving a hint of the gene being regulated by ethylene. ERE has already been associated with tomato fruit ripening gene E4, which responds to ethylene (Itzhaki et al., 1994). *In-silico* analysis identified many light responsive elements like G-Box, I-box and GT-1 confirming its dependence on light for its expression. This result was further supported by *GalUR* (D-Galactouronic acid reductase) promoter from strawberry, responsible for vitamin C synthesis and expressed during fruit ripening, in which all the light responsive elements were conserved and the gene expression was regulated by dark/light treatments (Agius et al., 2005). Although the *GalUR* promoter is from a non-climacteric fruit and *SlAMS* from a climacteric fruit, but these elements seem to be conserved between the two. Furthermore, these elements are important characteristic of light regulated genes (Terzaghi and Cashmore, 1995). An interesting motif was identified by PLACE in the promoter region of *SlAMS* called the "CGCG box" this is found in the promoters of those genes involved in ethylene signaling, abscisic acid singling, and light signal perception (Yang and Pooviah, 2002). One of the calmodulin binding protein, *AtSR1* (*Arabidopsis thaliana* signal-responsive) targets the "CGCG box" and modulates the gene during different environmental signals suggesting that the gene promoters possessing this box are regulated by calmoduling-binding/DNA binding proteins and are involved in multiple signal transduction pathway in plants (Yang and Pooviah, 2002). Another homolog of *AtSR1*, *NtER1* from tobacco is also a CaM binding protein which is an early ethylene-up-regulated gene giving evidence of Ca^{2+}

mediated signaling in ethylene action, in the genes with the "CGCG box" in their promoter (Yang and Pooviah, 2000). Furthermore the *NtER1* is also involved in plant senescence and death pointing towards the possible role of the gene in programmed cell death. The *SlAMS* promoter was also having GCC box, which is a hallmark of pathogenesis-related (PR) genes. The Pti4 (transcription factor) protein is known to bind to GCC box in pathogenesis-related (PR) genes and induce its expression (Chakravarthy et al., 2003). The GCC box can also sense ethylene and biotic stress, like infection with *Pseudomonas syringae* (Gu et al., 2000).

The existence of several stress responsive and ripening-regulated elements such as Myb, LTRE, ABRE, ERE, CGCG box, GCC box and G-box support the conclusion that the *SlAMS* gene is both biotic and abiotic stress-inducible and also a ripening-related gene. Similar kind of elements was found in *asr-1(abscisic acid responsive)* gene from melon involved in fruit ripening (Hong et al., 2002). Although many regulatory cis-elements are conserved among diverse set of unrelated genes as well as among the gene with different respiratory behavior, functional proofs are required to validate the promoter in different species and organs to be exploited for different purposes.