Available online at www.sciencedirect.com
ScienceDirect
Biochemical and Biophysical Research Communications 364 (2007) 1056–1061
BBRC
www.elsevier.com/locate/ybbrc

Structure and alternative splicing of a heat shock transcription factor gene, *MsHSF1*, in *Medicago sativa*

Zhi-shui He [a,b], Rong Xie [a,b], Hua-song Zou [a], Yan-zhang Wang [a], Jia-bi Zhu [a], Guan-qiao Yu [a,*]

[a] *National Laboratory of Plant Molecular Genetics, Institute of Plant Physiology & Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, 300 Fenglin Road, Shanghai 200032, PR China*
[b] *Graduate School of the Chinese Academy of Sciences, Beijing 100049, PR China*

Received 17 October 2007
Available online 30 October 2007

Abstract

Plant heat shock transcription factors (HSF) are highly complex. In this study, we identified an alfalfa HSF gene *MsHSF1* that is composed of four exons and three introns in the encoding region. The intron1–exon2–intron2–exon3–intron3 as an intervening sequence was inserted at the conserved position that separates the coding region for the DNA-binding domain by single intron in other known plant HSF genes. Alternative splicing of *MsHSF1* has generated five transcript isoforms. Spliced transcript *MsHSF1b* consisted of exon1 and exon4, encodes a class A1 HSF protein that can specifically bind to the heat shock elements *in vitro*. Other four spliced transcripts (*MsHSF1a-1* to *4*) consist of exon1, part of the intervening sequence and exon4. These transcripts carry the premature termination codon and are low-abundant. Apparently these transcripts are the targets of nonsense-mediated mRNA decay (NMD). These results provide new insight into roles in the expression regulation of plant HSF genes.

Keywords: Heat shock transcription factor; Alternative splicing; Nonsense-mediated mRNA decay; Root nodule

Plant heat shock transcription factors (HSFs) contain a DNA-binding domain (DBD), an oligomerization domain carrying two adjacent hydrophobic heptad repeats (HR-A/B), and a nuclear localization signal (NLS); in most cases they also have a nuclear export signal (NES) [1]. The DBD is the most conserved region in eukaryotic HSFs. The known plant HSFs contain a single intron of varying size immediately downstream of the helix-turn-helix motif (H2-T-H3) in the DBD, except for *MsHSFA4*, in which the intron is shifted six amino acid residues downstream from the conserved position [1–3]. A single intron in this conserved position is characteristic of plant HSFs [1,4].

Another common feature of many HSFs is alternative splicing [5]. In *Drosophila*, *Schistosoma mansoni*, and vertebrates, HSFs have alternative-spliced isoforms [5–12]. These alternative-spliced isoforms are generated by insertion or deletion of a fragment of amino acid sequence at the downstream of HR-A/B. The only example of alternative splicing inside the DBD is the splicing of *SmHSF* with retention of intron1 and 2 in *S. mansoni* [13]. Retention of intron2 containing the termination codon leads to decay of *SmHSF*. The position of intron2 in *SmHSF* is identical to that of the single intron of plant HSFs, suggesting that this conserved position may have been important in HSF gene evolution [7]. To date, for alternative splicing of plant HSFs, only *AtHSF3* (AT5g16820), *AtHSF6* (AT5g62020), a rice HSF(OsO3g53340) and an Arabidopsis HSF(AT5g03720) have been found by genomics analysis [14,15].

Here we report that *MsHSF1* contained an intervening sequence of intron1–exon2–intron2–exon3–intron3 inserted at the conserved position that separates the coding region for DBD by only one intron in other known plant

* Corresponding author. Fax: +86 21 54924015.
E-mail address: gqyu@sippe.ac.cn (G. Yu).

0006-291X/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.bbrc.2007.10.131

HSF genes. Splicing this intervening sequence generates multiple transcript isoforms with premature termination codon (PTC). Study of the organizations and functions of the intervening sequence at the conserved position enable further understanding of the expression regulation of plant HSFs genes.

Materials and methods

Plant materials. Plant growth conditions were described previously [16]. Briefly, seeds of alfalfa (*Medicago sativa* L. cv. Algonquin) were sterilized in 1% sodium hypochlorite, rinsed and germinated on 1% agar and then inoculated with *Sinorhizobium meliloti* wild-type strain Sm1021. The seedlings grew in a nitrogen-free medium under controlled environmental conditions (16 h day/8 h night cycle, 24 °C/20 °C day/night temperature, and 70% RH). 15-day-old nodules, stems, leaves, uninoculated roots and mature flower were collected. All tissues were quick-frozen in liquid nitrogen, and stored at −70 °C.

Preparation of genomic DNA and cDNA. Total RNA was extracted from various tissues using the TRIZOL reagent according to the manufacturer's instructions (Invitrogen) and treated with RNase–free DNase I (Takara). The quality and concentration of purified total RNA were determined by UV spectrophotometer and formaldehyde gel electrophoresis. The first-strand cDNA was prepared from total RNA using oligo(dT)18 and the SuperScript III (Invitrogen) following the manufacturer's protocol. Genomic DNA was isolated from leaves according to the CTAB protocol [17].

PCR amplification, cloning, and sequencing. All PCRs were carried out in a reaction volume of 50 μL with 1× buffer (for KOD-Plus, Toyobo), 0.2 mM dNTP, 1 mM $MgSO_4$, 0.4 μM primers and 1.0 unit KOD-Plus DNA polymerase (Toyobo). The PCR products were cloned to the pMD18-T vector (Takara), sequenced or analyzed by electrophoresis on 1.0–2.0% agarose gel, and stained with ethidium bromide. Each clone was repeated a minimum of three times.

Primers 5′-CCCCTTTTCTCCCTTCAAATTC-3′ (F1) and 5′-AAAGCTTCTCGGTGAGCAAATG-3′ (R1) and internal primers 5′-CATTCAAAAACAAGAATTCTCCG-3′ (F2) and 5′-GGGTCAGG GATTACCTTGGA-3′ (R2) were used to isolate the *MsHSF1* genomic DNA and *MsHSF1b* cDNA by nested PCR. The template was denatured at 94 °C for 2 min, followed by 35 cycles of 94 °C, 30 s; 56 °C, 30 s; and 68 °C, 5 or 2 min.

To identify the alternative-splicing patterns by reverse transcriptase-PCR (RT-PCR), two sets of primers were used. The first set of primers corresponded to exon1, 5′-AGGGACCTTTTGCCGAAGTTCTTCA-3′ (AF1) and to exon4, 5′-CCAACGATCTGGGTCAACCTTCC-3′ (AR1). Amplification was performed at 94 °C for 2 min, followed by 35 cycles of 94 °C, 30 s; 60 °C, 30 s; and 68 °C, 30 s. The other set of primers corresponded to exon2, 5′-GGAGACAGGACAACCTTTGCCGTAC-3′ (AF2) and to exon3, 5′-AGCTTTCTTTGTCAAATCTGCCTCT-3′ (AR2). Amplification was performed at 94 °C for 2 min, followed by 30–35 cycles of 94 °C, 30 s; 54 °C, 30 s; and 68 °C, 40 s.

The *Msc27* [18] was used as an internal control for RT-PCR with primers 5′-GGAATGTTGTGGGAGGTTGA-3′ and 5′-AAAGAATT GAAGGTCCTTGAGC-3′. Reaction conditions were 94 °C for 2 min followed by 30 cycles of 94 °C, 30 s; 58 °C, 30 s; and 68 °C, 40 s.

Expression and purification of recombinant MsHSF1b in Escherichia coli. The coding region of *MsHSF1b* was cloned in-frame into the pET28b vector (Novagen) between the SacI and XhoI sites. The integrity of the recombinant plasmid was verified by DNA sequencing. The recombinant plasmid was transformed into the BL21 (DE3). Transformed cells were grown at 37 °C with 1 mM IPTG for 3 h. His-tagged MsHSF1b was purified by affinity chromatography through Ni–NTA resin followed by elution with 200 mM imidazole. The protein was desalted and concentrated using Amicon Ultra-15 filter devices (30 000 MWCO, Millipore) with the buffer (20 mM Tris–HCl pH 7.9, 50 mM NaCl, 1 mM $MgCl_2$, 0.5 mM EDTA, 1 mM DTT, 10% glycerol). The protein concentration was then determined by Bradford protein assay, and aliquots of purified protein were frozen in liquid nitrogen and stored at −70 °C.

Electrophoretic mobility shift assay. The mobility shift assay was carried out as described by Hübel and Schöffl [19] with minor modifications. Briefly, 1 μg purified MsHSF1b was incubated with 2.5 ng ^{32}P-labeled heat shock elements (HSE) oligonucleotides in 20 μL binding buffer (20 mM Tris–HCl pH 7.9, 50 mM NaCl, 10 mM $MgCl_2$, 0.5 mM EDTA, 1 mM DTT, 10% glycerol, and 1 μg salmon DNA) at room temperature for 20 min. The mixtures were electrophoresed on a 4% polyacrylamide gel at 4 °C, then autoradiographed. The oligonucleotides [19] were 3′-end-labeled by filling-in with α-^{32}P-dCTP using exo-free Klenow fragment (Takara).

HSE

```
5'  tcgaggatcctaGAAgcTTCcaGAAgcTTCtaGAAgcagat
          ||||||||||||||||||||||||||||||||||||
          ctaggatCTTcgAAGgtCTTcgAAGatCTTcgtctagagct 5'
```

HSEm

```
5'  tcgaggatcctaTAAgcTTAcaTAAgcTTAtaTAAgcagat
          ||||||||||||||||||||||||||||||||||||
          ctaggatATTcgAATgtATTcgAATatATTcgtctagagct 5'
```

Phylogenetic analysis. Full-length amino acid sequences were aligned using the CLUSTAL W 1.83 software [20]. Phylogenetic analysis was conducted using MEGA version 3.1 [21]. Distance matrices used the PAM-Dayhoff model. A phylogenetic tree was constructed using the neighbor-joining method with the ScHSF [22] as the outgroup. Bootstrap analyses of 1000 resampling replicates were made to test for the statistical significance of nodes. To simplify the analysis, only some representative HSF protein sequences were used.

Results

MsHSF1 contains four exons and three introns in the encoding region

A nodule-enhanced cDNA clone RX74 (*MsHSF1a-2*) (GenBank Accession No.: CB858136) whose corresponding gene presumably encodes a HSF has been identified [16]. Sequence comparison with *Medicago truncatula* genomic DNA (*MtHSF1g*) (GenBank Accession No.: AC087771) revealed that the coding region of *MtHSF1g* was composed of four exons interrupted by three introns; and *MsHSF1a-2* consists of these four exons, and the deduced amino acid sequence contained at least one PTC in the reading frames.

To examine the organization of the *MsHSF1a-2*-corresponding gene (*MsHSF1*) we cloned *MsHSF1* from alfalfa leaf (Table 1). The detailed comparisons of *MsHSF1g-1* (GenBank Accession No.: DQ907239) with *MtHSF1g* showed that *MsHSF1* was also composed of four exons and three introns in the encoding region (Fig. 1A and Table 1). The lengths of the corresponding exons were identical for each gene in the encoding region, but that of corresponding introns differed, except for intron2 (Table 1). The AT content of all introns was between 65.0% and 73.7% (Table 1), similar to other plant introns [23–25]. The AT content of introns in dicotyledons (≈60%) is necessary for efficient splicing [26]. The sequences at exon–intron junctions all follow the GT-AG rule [27].