Genomics 97 (2011) 166–172

Contents lists available at ScienceDirect

Genomics

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

A common sequence motif involved in selection of transcription start sites of Arabidopsis and budding yeast tRNA genes

Yasushi Yukawa [a,*], Giorgio Dieci [b], Mircko Alzapiedi [b], Asako Hiraga [a], Katsuaki Hirai [a], Yoshiharu Y. Yamamoto [c], Masahiro Sugiura [a]

[a] Graduate School of Natural Sciences, Nagoya City University, 467-8501 Nagoya, Japan
[b] Dipartimento di Biochimica e Biologia Molecolare, Università degli Studi di Parma, 43100 Parma, Italy
[c] Faculty of Applied Biological Sciences, Gifu University, Gifu 501-1193, Japan

ARTICLE INFO

Article history:
Received 1 September 2010
Accepted 2 December 2010
Available online 13 December 2010

Keywords:
Transcription start site
tRNA gene
Pol III
Arabidopsis
Budding yeast

ABSTRACT

The transcription start site (TSS) is useful to predict gene and to understand transcription initiation. Although vast data on mRNA TSSs are available, little is known about tRNA genes because of rapid processing. Using a tobacco *in vitro* transcription system under conditions of impaired 5′ end processing, TSSs were determined for 64 Arabidopsis tRNA genes. This analysis revealed multiple TSSs distributed in a region from 10 to 2 bp upstream of the mature tRNA coding sequence (−10 to −2). We also analyzed 31 *Saccharomyces cerevisiae* tRNA genes that showed a smaller number but a broader distribution (−13 to −1) of TSSs. In both cases, transcription was initiated preferentially at adenosine, and a common 'TCAACA' sequence was found spanning the TSSs. In plant, this motif caused multiple TSSs to converge at one site and enhanced transcription. The TATA-like sequence upstream of Arabidopsis tRNA genes also contributed to TSS selection.

1. Introduction

The nucleus in plant cells possesses three major RNA polymerases (Pol I, Pol II, and Pol III). Pol I transcribes rRNA gene clusters, and Pol II synthesizes mRNAs and small RNAs including snRNAs, snoRNAs, and microRNAs. Pol III produces additional small RNAs, such as tRNAs, 5S rRNA, U6 snRNA, and U3 snoRNA (animal and yeast U3 snoRNAs are synthesized by Pol II), 7SL RNA, MRP RNA (yeast MRP RNA is made by Pol II), SINE RNAs, and novel classes of non-coding RNAs [1]. However, genes for RNase P RNA and telomerase RNA, which are Pol III-dependent in animals and yeast, have not yet been confirmed in plants [2–4]. Recently, a fourth and fifth RNA polymerase (Pol IV and Pol V, previously known as Pol IVa and Pol IVb) were reported to be present in Arabidopsis and involved in RNAi-mediated DNA methylation and gene silencing [5–11]. Although Pol II has received the most attention, Pol I and Pol III are responsible for over 80% of total RNA synthesis in growing cells [12].

To study transcriptional processes, it is necessary to define transcribed regions or at least transcription start sites (TSS). For protein-coding genes, a large number of TSSs have been accumulated by methods based on the 5′ cap structures of transcripts, for example, full-length cDNAs [13–15] and cap analysis of gene expression (CAGE) [16]. The TSSs of Pol III-dependent genes cannot be determined by the above methods, and therefore transcripts have generally been excluded from traditional transcriptome analyses. To investigate how Pol III recognizes small RNA genes scattered over the genome and starts transcription at the proper sites, information on TSSs is required. However, it is generally difficult to detect Pol III-dependent primary transcripts, for example pre-tRNAs, due to rapid processing *in vivo*.

We have developed an *in vitro* transcription system from cultured tobacco cells and studied the transcription of plant Pol III-dependent genes for tRNA, 7SL RNA, 5S rRNA, and SINE retroposons [17–23]. We recently assayed *in vitro* transcription of 124 annotated tRNA genes in Arabidopsis chromosome I, and estimated that 16% or more of the annotated tRNA genes are not functional [24]. Our *in vitro* system supports not only tRNA transcription but also 5′- and 3′-end processing, nucleotide modification, and splicing of intron-containing pre-tRNAs [19]. The optimal Mg^{2+} concentration differs for each step, and transcription only occurs at a low Mg^{2+} concentration so that pre-tRNAs can accumulate. Unlike Pol II transcription termination (e.g. [25]), Pol III-dependent genes contain a clear termination signal, the T-stretch present downstream from tRNA genes, which is necessary for robust re-initiation of transcription [26].

Using the *in vitro* transcription system under a low Mg^{2+} concentration and with no reinitiation, we have now determined the TSSs of 64 tRNA genes from *A. thaliana*. Similarly, we also determined the TSSs of 31 tRNA genes from the budding yeast, *S.*

* Corresponding author. Graduate School of Natural Sciences, Nagoya City University, Yamanohata 1, Mizuho-cho, Mizuho-ku, 467-8501 Nagoya, Japan. Fax: +81 52 872 6021.
E-mail address: yyuk@nsc.nagoya-cu.ac.jp (Y. Yukawa).

0888-7543/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2010.12.001

cerevisiae, using a homologous *in vitro* transcription system. We found that both systems preferentially initiate tRNA transcription with an adenosine residue. A common motif, typically TCAACA, is present spanning the TSSs of many Arabidopsis and budding yeast tRNA genes, and mutational analysis indicated that this motif directs TSS selection. Furthermore, the TATA-like motif and A-box are also involved in TSS selection. Our data should be useful for predicting Pol III-dependent genes encoding novel non-coding RNAs.

2. Results and discussion

2.1. Determination of transcription start sites (TSSs)

We previously assayed *in vitro* the transcription activities of 124 tRNA genes in Arabidopsis chromosome I [24]. The 64 genes with higher transcription activities were selected to determine their TSSs. Fig. 1 shows our assay procedures. To prepare tRNA gene templates, 5′-primers (21~30 nt) were designed ca. 400 bp upstream from the 5′-ends of tRNA coding regions and the 3′-primers corresponded to the 3′ tRNA regions (28 nt) with the M13 reverse sequence (19 nt) as a tag. *In vitro* transcription from linear templates with tobacco nuclear extracts were performed with 1 mM Mg^{2+}, because only transcription proceeds efficiently while 5′ processing is inhibited at this concentration [19,27]. The obtained linear templates have no terminator sequence (T-stretch), so that transcription reinitiation is strongly disfavored and only a single round of tRNA gene transcription can occur under our reaction conditions (data not shown). Restriction to the first round of transcription renders this method useful to investigate how Pol III recognizes the site at which it first initiates transcription, without any background from subsequent cycles (as reinitiation TSS might in principle differ from first round TSS). *In vitro* synthesized pre-tRNA were subjected to primer extension using a FITC- or ^{32}P-labeled 3′-reverse tag primer.

Arabidopsis TSS data thus obtained are summarized in Fig. 2A. Thirteen tRNA genes possess single TSSs at positions -6 to -4, while other genes have 2 to 6 TSSs at positions -10 to -2; major TSSs are indicated by big dots and weak TSSs by small dots. The TSSs of 26 tRNA genes of *S. cerevisiae* were also assayed by primer extension analysis of pre-tRNAs synthesized *in vitro* in a purified transcription system in which 5′-processing does not occur appreciably [28]. The results of this analysis are shown in Fig. 2B, which also shows information on the TSSs of 5 tRNA genes determined in previous studies [Gln_ttg 133 = tQ{TTG}HR, Tyr_gta 90 = tY{GTA}FR2, Tyr_gta 155 = tY{GTA}JR, Leu_caa 18 = tL{CAA}CL: [29]; Ile_tat 199 = tI{TAT}LR1: [28]]. Fourteen genes possess single TSSs and the remaining 17 genes have 2 to 4 TSSs, indicating that single TSSs are more frequent in yeast than in Arabidopsis. No TSS was found in tRNA coding regions, indicating that Pol III first recognizes several nucleotides upstream from the 5′-end of mature tRNA regions and starts transcription. This observation suggested the existence of a sequence element(s) supporting TSS recognition together with internal promoter boxes. For 26 of the 31 yeast tRNA genes, the experimentally determined TSS matched the one predicted by Kassavetis (supplement to [30]).

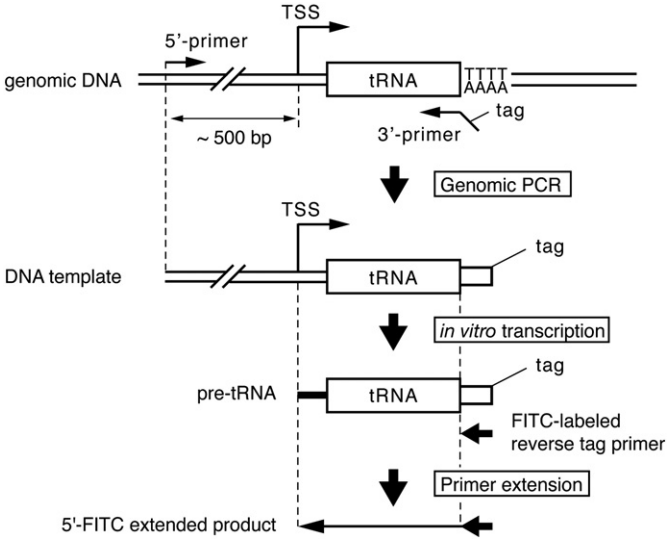

Fig. 1. Schematic representation of constructs for the determination of transcription start sites (TSSs) of tRNA genes *in vitro*. Linear DNA templates were prepared by genomic PCR using 5′-primers (straight arrows) and M13 reverse sequence tagged 3′-primers (bent arrows). *In vitro* transcripts (pre-tRNAs) were subjected to reverse transcription with a FITC-labeled reverse tag primer.

2.2. Common sequence surrounding TSSs

As shown in Fig. 3A, adenosine residues are preferentially selected as TSSs both for Arabidopsis (66%) and yeast (85%) genes, and TSSs are rare at guanosine (Arabidopsis: 9%, yeast: 1.6%) and thymidine (Arabidopsis: 4%, yeast: 5%) residues. All 13 single TSSs in Arabidopsis are A residues, and 13 out of the 14 single TSSs in yeast are also A residues. This observation is surprising because prokaryotic RNA polymerases and Pol II use both A and G as TSSs, and phage T7 RNA polymerase starts transcription strictly at G. In the case of plant 5S rRNA, U6 snRNA and 7SL RNA genes, Pol III transcription initiates at G [21,31,32], thus suggesting that selection against G's observed at tRNA genes is not merely due to sequence preference by Pol III, but possibly to a tRNA-specific mechanism of TSS selection influenced by Pol III transcription factors.

A remarkable difference between Arabidopsis and yeast TSSs is the distribution of their positions (Fig. 3B). In Arabidopsis, TSSs show a single peak distribution at position -5 (numbering is with respect to the 5′-end of mature tRNA). In yeast, a broader TSSs distribution is found, with three peaks observed at -10 (dominant peak), -5 and -2. As a consequence, plant pre-tRNAs contain 5′ leader sequences that are generally ~5 nt shorter than yeast pre-tRNAs (cf. [33]).

tRNA genes include two highly conserved promoter elements (A-box and B-box) within transcribed regions. The A-box consensus is RTRGYNYAGTGGT (R: purine, Y: pyrimidine, N: any base, [34]) and the 3′ side GG dinucleotides are conserved in all organisms (Fig. 3C and S1). With respect to the GG position in the A-box, all Arabidopsis tRNA genes tested here are classified into two types: the (17•18) type and (18•19) type, where the GG is located at +17/+18 and +18/+19 positions, respectively. Fig. 3D shows TSS distributions between (17•18) type and (18•19) type genes, and illustrates a clear difference of peaks at −6 and −5 positions, respectively. Most TSSs emerge 22 nt upstream from the GG dinucleotides. This observation suggests that the TSS is influenced by the GG position.

To search for short conserved sequences surrounding the identified TSSs, all of the sixty-four 5′ flanking sequences of Arabidopsis tRNA genes (from −50 to +1, see Fig. 2) were subjected to the Multiple EM for Motif Elicitation program (MEME, http://meme.sdsc.edu/, [35]), and an evident motif was detected. The basic form of the motif is TCȦȦCA (dots indicate TSSs), but it can diverge as (T/A)CȦ(A/T)(C/A)(A/T), as shown in Fig. 3E. A similar motif is also found in the 5′ flanking regions of yeast tRNA genes (see Fig. 2B, indicated by underlines, and Fig. 3E). In yeast, the motif partially overlaps with TSS-associated conserved sequence elements identified by previous computational analyses [30,36,37]. We named the conserved sequence the TCAACA motif. However, several tRNA