Biochemical and Biophysical Research Communications 423 (2012) 289–294

Contents lists available at SciVerse ScienceDirect

Biochemical and Biophysical Research Communications

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

Mechanistic insights into human pre-mRNA splicing of human ultra-short introns: Potential unusual mechanism identifies G-rich introns

Noriko Sasaki-Haraguchi [a], Makoto K. Shimada [a], Ichiro Taniguchi [b], Mutsuhito Ohno [b], Akila Mayeda [a,*]

[a] Division of Gene Expression Mechanism, Institute for Comprehensive Medical Science, Fujita Health University, Toyoake, Aichi 470-1192, Japan
[b] Institute for Virus Research, Kyoto University, Kyoto 606-8507, Japan

ARTICLE INFO

Article history:
Received 15 May 2012
Available online 26 May 2012

Keywords:
Pre-mRNA splicing
Ultra-short intron
Intronic splicing enhancer
G-rich sequence
U snRNA
Spliceostatin A

ABSTRACT

It is unknown how very short introns (<65 nt; termed 'ultra-short' introns) could be spliced in a massive spliceosome (>2.7 MDa) without steric hindrance. By screening an annotated human transcriptome database (H-InvDB), we identified three model ultra-short introns: the 56-nt intron in the *HNRNPH1* (hnRNP H1) gene, the 49-nt intron in the *NDOR1* (NADPH dependent diflavin oxidoreductase 1) gene, and the 43-nt intron in the *ESRP2* (epithelial splicing regulatory protein 2) gene. We verified that these endogenous ultra-short introns are spliced, and also recapitulated this in cultured cells transfected with the corresponding mini-genes. The splicing of these ultra-short introns was repressed by a splicing inhibitor, spliceostatin A, suggesting that SF3b (a U2 snRNP component) is involved in their splicing processes. The 56-nt intron containing a pyrimidine-rich tract was spliced out in a lariat form, and this splicing was inhibited by the disruption of U1, U2, or U4 snRNA. In contrast, the 49- and 43-nt introns were purine-rich overall without any pyrimidine-rich tract, and these lariat RNAs were not detectable. Remarkably, shared G-rich intronic sequences in the 49- and 43-nt introns were required for their splicing, suggesting that these ultra-short introns may recruit a novel auxiliary splicing mechanism linked to G-rich intronic splicing enhancers.

1. Introduction

The length distributions of vertebrate pre-mRNA introns have a bimodal pattern, e.g., a narrow distribution peaking around ~100 nucleotides (nt) ('short introns') and a broad distribution peaking around ~2000 nt ('long introns') in humans [1,2]. Although intron-definition and exon-definition models have been proposed for the splicing of short and long introns, respectively [3], we do not know much about the exact splicing mechanisms for introns in extra-range below and above of the bimodal distribution. In fact, the mechanism of pre-mRNA spicing has been studied with model pre-mRNAs containing single short introns (100–250 nt), which have been spliced efficiently *in vivo* and *in vitro* (reviewed in [4,5]).

Here, we focus on very short introns in the range below the narrow distribution of the first mode, i.e., less than 65 nt in length, which we designated 'ultra-short introns' (Shimada et al., manuscript in preparation). Early studies showed that a minimum intron length of 78–80 nt is necessary for their splicing [6,7]. However, the real minimum length cannot be inferred from these data because these tested introns were arbitrarily shortened from natural introns, which would eliminate important *cis*-acting elements [8,9]. We carefully screened human genes containing functional ultra-short introns and the splicing of identified three model ultra-short introns (43, 49, and 56 nt) was verified *in vivo* and *in vitro*.

Human nuclear pre-mRNA splicing involves dynamic stepwise reactions in a huge spliceosome, which includes five kinds of U snRNPs and ~170 proteins (reviewed in [5,10]). Essential *cis*-reactive elements in the pre-mRNA; the 5′ splice site, the branch-point sequence, and the 3′ splice site, are simultaneously bound by corresponding *trans*-acting factors, U1 snRNP, U2 snRNP, and $U2AF^{65}$/$U2AF^{35}$, respectively, leading to the initial ATP-dependent formation of the spliceosomal A complex (~2.5 MDa) [11]. Structural analysis by electron microscopy has estimated that the A complex has an asymmetric globular shape, with dimensions of ~26 × 20 × 19.5 nm [11]. Thus, the A complex fully occupies the length of an 85–113-nt linearized RNA (as 1-nt RNA = 0.23 nm; reference [12]), which is about twofold longer than the ultra-short introns we tested (43–56 nt). Therefore, we asked how these three essential splicing signals can be recognized by the corresponding essential factors without steric hindrance.

2. Materials and methods

Full descriptions are provided in the Supplementary Materials and methods.

Abbreviations: SSA, spliceostatin A; PAGE, polyacrylamide gel electrophoresis; ORF, open reading frame; ISE, intronic splicing enhancer; ESE, exonic splicing enhancer.

* Corresponding author. Fax: +81 562 93 8834.
E-mail address: mayeda@fujita-hu.ac.jp (A. Mayeda).

0006-291X/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.bbrc.2012.05.112

2.1. Construction of expression plasmids

The expression plasmids for transient transfections of cultured cells were constructed by subcloning the corresponding PCR fragments from the *HNRNPH1, NDOR1, and ESRP2* genes, together with PCR fragments from the pBI-EGFP plasmid, into the pCAGGS vector [13]. These amplified products were subcloned into the pGEM-T Easy vector with the MS2-binding domain fragment from the AdML-M3 plasmid [14,15] for the templates of *in vitro* transcription. PCR amplifications were performed with Blend Taq DNA polymerase and the corresponding DNA primers.

2.2. Transfection of cells and RT–PCR analysis

The indicated human cell lines were transiently transfected with the indicated expression plasmids using Lipofectamine LTX. The transfected cells were incubated for a total of 22 h. Spliceostatin A (SSA) [16] was added 16 h after transfection. Their total cellular RNA was extracted with the PureLink RNA Mini Kit, digested with DNase I, and reverse transcribed using the PrimeScript reverse transcriptase with an oligo-dT primer or specific primers. All the PCR products were analyzed by 5% polyacrylamide gel electrophoresis (PAGE).

2.3. In vitro splicing assays and detection of lariat introns

To prepare [^{32}P] UTP-labeled hnRNP H1, NDOR1, ESRP2, and β-globin pre-mRNAs, *in vitro* transcription with T7 or SP6 RNA polymerase was performed as described previously [17]. The *in vitro* splicing assays were performed as described [17], with the indicated modifications. The selective detection of lariat products by RNase R digestion was performed as described previously [18]. The digested products were analyzed by denaturing 7% PAGE, and visualized using a Bio-imaging Analyzer (Fujix BAS1000, Fujifilm) as described [17].

2.4. Disruption of U snRNAs and splicing assays in Xenopus oocytes

The microinjection of RNA into *Xenopus* oocytes and the preparation of the RNA were performed as described previously[19]. Antisense oligonucleotide directed against U1, U2, or U4 snRNA [20,21] was injected into the *Xenopus* oocyte cytoplasm to disrupt these snRNAs by the endogenous RNase H activity, as described previously [22].

3. Results

3.1. Human ultra-short introns were screened and identified

Using the intronic sequences from the human transcript database (H-InvDB, version 6.0) and the human genome database (NCBI build 36), we derived the pattern of the length distribution of the human introns that are shorter than 945 nt (Fig. 1A, left panel). We observed that the number of human introns increases drastically from 65 nt toward the first mode of the distribution at 83 nt, which is essentially consistent with previous reports [1,2]. Most of these short introns could be spliced by the well-known conventional mechanism because the mechanism was established using short introns in this range, either *in vitro* or in transfected cells. However, we assumed that a considerable number of candidate introns shorter than the 65-nt threshold would not be spliced by the known conventional mechanism, and thus categorized them as 'ultra-short introns' (Fig. 1A, right panel).

There are 190 counts of ultra-short candidate introns shorter than 57 nt, with the terminal GT and AG nucleotides. By removing the candidate sequences that included possible alignment mistakes and varying numbers of tandem repeats, we identified 33 candidate ultra-short introns (Supplementary Table S1). We further screened these ultra-short introns based on their locations in an open reading frame (ORF), the functional identities of their host protein-coding genes, and their conservation in primates, such as the chimpanzee and macaque (Supplementary Materials and methods). In this way, we selected three representative model ultra-short introns: the sixth intron (56 nt) in the hnRNP H1 (*HNRNPH1*) gene (an annotated transcript information in HIT000192494), the twelfth intron (49 nt) in the NADPH-dependent diflavin oxidoreductase-1 (*NDOR1*) gene (HIT000009363), and the sixth intron (43 nt) in the epithelial splicing regulatory protein-2 (*ESRP2*) gene (HIT000008845) (Fig. 1B). There was no specific bias in the positions of the ultra-short introns in their host genes. Remarkably, the shorter two ultra-short introns (49 and 43 nt) do not possess any apparent consensus sequence for branch points, (C/T)TNA(C/T) followed by polypyrimidine tracts, that have been reported in human genes [23].

3.2. Selected human ultra-short introns were spliced in living cells

We checked whether these endogenous ultra-short introns are actually spliced in cultured cells by RT–PCR. The hnRNP H1 pre-mRNA, containing the 56-nt intron, was efficiently spliced. We found distal alternative 3′ splice sites in the 49- and 43-nt introns of the NDOR1 and ESRP2 pre-mRNAs, which generate 76- and 73-nt introns, respectively. These distal 3′ splice sites were used more efficiently than the proximal 3′ splice sites (Fig. 1B, panels for NDOR1 and ESRP2). Especially, a specific inner reverse primer was required to clearly detect the proximal 49-nt NDOR1 intron (Fig. 1B, lower panel of NDOR1). These data confirm that the three model ultra-short introns in the endogenous hnRNP H1, NDOR1, and ESRP2 pre-mRNAs are indeed spliced in cultured cells.

3.3. Splicing factor SF3b is involved in splicing of ultra-short introns

Because the ultra-short introns are too short to be assembled with the known authentic spliceosome, which includes essential factors such as the U1 snRNP, U2 snRNP, and U2AF^{65}/U2AF^{35} in the early step [4], we assumed that their splicing is catalyzed in a more compact spliceosome that lacks the known essential factors.

To test this assumption, we first used the splicing inhibitor, SSA, which interferes with splicing through the specific binding to SF3b (a component of the U2 snRNP) [16]. Human cells were transfected with mini-gene encoding the 56-nt (hnRNP H1), 49-nt (NDOR1), or 43-nt (ESRP2) intron with or without SSA in the culture medium (Fig. 2A). We found that SSA significantly inhibited the splicing of these ultra-short introns, as well as the splicing *via* the distal alternative 3′ splice sites of NDOR1 and ESRP2. These data demonstrate that SF3b is not a dispensable splicing factor for these ultra-short introns.

3.4. U1, U2, and U4 snRNPs are required to splice the 56-nt intron

Using the established *Xenopus* oocyte microinjection system, we next tested whether the essential U snRNPs are required for splicing of these ultra-short introns. Each ^{32}P-labeled pre-mRNA containing the 56-nt (hnRNP H1), 49-nt (NDOR1), or 43-nt (ESRP2) intron was microinjected into oocytes of *Xenopus laevis*, together with the control 231-nt adenovirus major late intron (AdML). The 56- and 231-nt introns were efficiently spliced (Fig. 2B, lanes 1, 3), whereas the 49- and 43-nt introns were not spliced (data not shown). Therefore, we could only use the hnRNP H1 pre-mRNA