Biochimica et Biophysica Acta 1789 (2009) 624–633

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

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

Review

Spliceosome structure: Piece by piece

Dustin B. Ritchie, Matthew J. Schellenberg, Andrew M. MacMillan *

Department of Biochemistry, University of Alberta, Edmonton, Alberta, Canada T6G 2H7

ARTICLE INFO

Article history:
Received 12 May 2009
Received in revised form 22 August 2009
Accepted 27 August 2009
Available online 3 September 2009

Keywords:
Spliceosome
Structure
X-ray
NMR
EM

ABSTRACT

Processing of pre-mRNAs by RNA splicing is an essential step in the maturation of protein coding RNAs in eukaryotes. Structural studies of the cellular splicing machinery, the spliceosome, are a major challenge in structural biology due to the size and complexity of the splicing ensemble. Specifically, the structural details of splice site recognition and the architecture of the spliceosome active site are poorly understood. X-ray and NMR techniques have been successfully used to address these questions defining the structure of individual domains, isolated splicing proteins, spliceosomal RNA fragments and recently the U1 snRNP multi-protein·RNA complex. These results combined with extant biochemical and genetic data have yielded important insights as well as posing fresh questions with respect to the regulation and mechanism of this critical gene regulatory process.

1. Introduction

The split gene structure of eukaryotes, in which protein-coding exon sequences are separated by non-coding intron sequences, is replicated at the level of the transcribed pre-messenger RNA (pre-mRNA). Intron excision and exon ligation occur concurrently in a process referred to as pre-mRNA splicing which is both a fundamental gene regulatory mechanism and a source of proteome diversity in higher eukaryotes [1–3].

The chemistry of pre-mRNA splicing involves two sequential transesterification reactions (Fig. 1). In the first step, the 2′ hydroxyl of a conserved adenosine within the intron carries out a nucleophilic attack at the 5′ splice site to generate a free 5′ exon and a cyclic (lariat) intermediate containing a 2′–5′ phosphodiester branch. Attack of the free 5′ exon at the 3′ splice site then yields ligated exons and the lariat intron product (Fig. 1A and B). These two steps of splicing are catalyzed by a massive ribonucleoprotein (RNP) complex referred to as the spliceosome that is reminiscent in terms of its large size and RNA·protein composition with the ribosome [1–3]. Indeed, the advancement of our understanding of the ribosome by high resolution structure determination [4,5] suggests that a similar structural analysis of the spliceosome is a worthwhile goal. Such an achievement would greatly enhance our understanding of splicing on multiple levels including both the regulation and basic chemical mechanism of pre-mRNA splicing. Complicating this endeavour is the complexity of the splicing machinery not only in terms of the sheer number of its constituents but also the nature of their interaction and dynamic association with each other and pre-mRNA substrate.

2. Spliceosome assembly and rearrangement

The spliceosome includes the U1, U2, and U4/U6·U5 snRNPs (small nuclear ribonucleoprotein particles) each containing a unique RNA and associated proteins. A central organizing feature of each snRNP is the set of seven Sm core proteins (LSm for U6 snRNP) that recognize the Sm binding site on each snRNA [6]. Estimates of the number of spliceosomal proteins vary but mass spectrometric analysis of affinity purified spliceosomes have suggested that upwards of 300 polypeptides may be associated with the splicing machinery [7–9]. Assembly of the spliceosome from smaller subunits is an ATP dependent process templated by the pre-mRNA substrate and is directed by conserved sequences at and proximal to the splice sites within the intron (Fig. 1C).

In the canonical pathway, discrete steps of spliceosome assembly through E (CC in yeast), A, B, and C complexes have been characterized [10]. Commitment of a pre-mRNA to splicing involves the ATP independent formation of the early (E) or commitment complex on the RNA substrate. This complex includes U1 snRNP, tightly associated with the 5′ splice site, as well as non-snRNP protein factors. In metazoans, these include the heterodimer U2AF, containing large and small subunits, which binds to the polypyrimidine tract and 3′ splice site and SF1 which recognizes the branch region [11,12]. The A complex is formed in part by the stable, ATP dependent, association of U2 snRNP with the pre-mRNA; a duplex formed between U2 snRNA and the pre-mRNA branch region bulges out the branch adenosine specifying it as the nucleophile for the first transesterification [13]. Association of the U4/U6·U5 tri-snRNP with the A complex produces B complex which undergoes a series of rearrangements to yield C complex, the mature spliceosome. These rearrangements include displacement of U1 snRNP at the 5′ splice site by U6 snRNP, the disruption of U4/U6 snRNA base-pairing, and the formation of a U2/U6 snRNA structure which is believed to form the active site of the spliceosome [2,3].

* Corresponding author.
E-mail address: andrew.macmillan@ualberta.ca (A.M. MacMillan).

1874-9399/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.bbagrm.2009.08.010

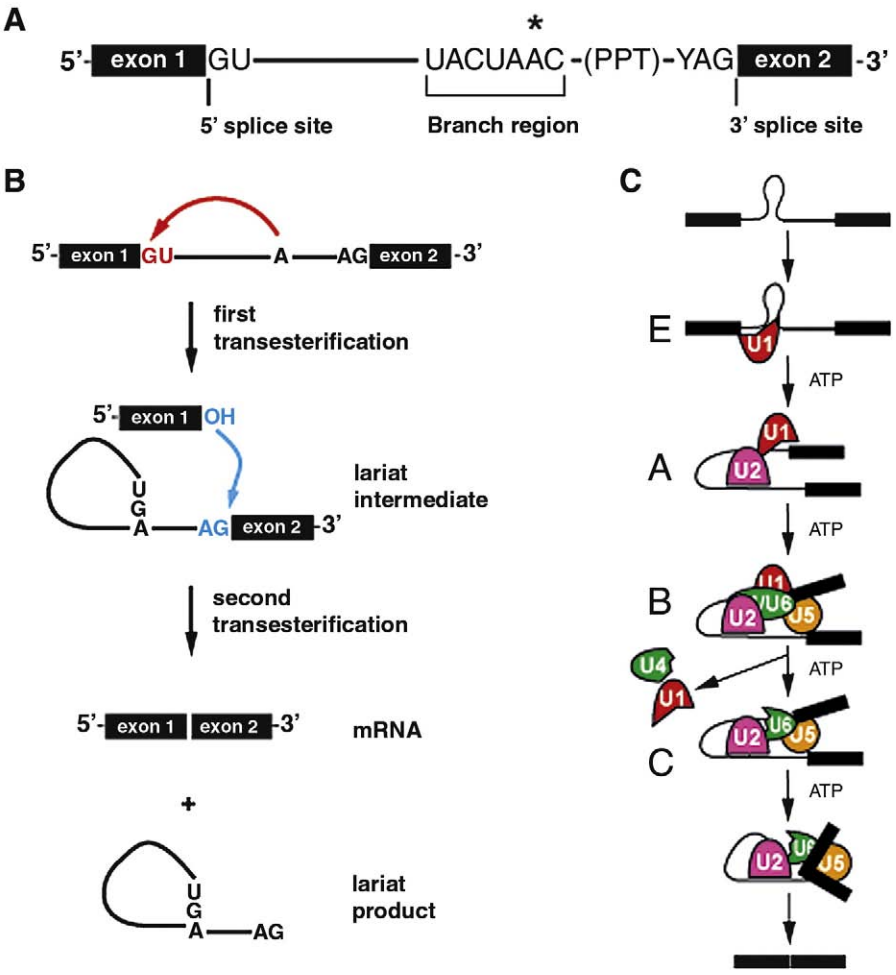

Fig. 1. Splicing of pre-mRNA by the spliceosome. (A) Intron structure highlighting conserved sequences at the 5′ and 3′ splice sites, the optimal branch sequence, and metazoan polypyrimidine tract. The preferred branch adenosine is indicated (*). (B) Sequential transesterification reactions catalyzed by the spliceosome. (C) Stepwise assembly of the spliceosome on a pre-mRNA showing sequential association of U1, U2, U4/U6·U5 snRNPs with the intron. Shown are the formation of the E (commitment), A, and B complexes through to the mature spliceosome in C complex.

3. Approaches to characterization of spliceosomal structure

Over and above non-trivial issues of abundance and purification, the dynamic nature of the spliceosome and the substrate dependence of its assembly have posed major difficulties in approaching structural studies. Current high resolution structural analysis of the spliceosome has relied primarily on a dissection of the complex into functionally important subunits amenable to analysis by X-ray or NMR (Table 1). With one exception, structural descriptions of higher order complexes have been restricted to EM studies as outlined here (Table 2).

This review highlights advances in our understanding of spliceosomal structure with an emphasis on what has been learned with respect to splice site recognition during the dynamic process of spliceosome assembly as well as the catalytic components at the heart of the splicing machinery.

4. Early recognition of pre-mRNA by spliceosomal factors: the E complex

4.1. Polypyrimidine tract recognition by U2AF

The U2 auxiliary factor (U2AF) is a heterodimer consisting of 55 and 35 kDa subunits. The large subunit (U2AF65) binds the polypyrimidine tract at the 3′ splice site, interacts with other E complex spliceosomal proteins [14] and, later, the U2 snRNP component SF3b155 [15]. U2AF65 contains an N-terminal RS domain followed by three regions originally described as RRMs (RNA recognition motifs). However, only the first two of these domains are involved in RNA binding; the third represents a general protein interaction domain referred to as a UHM (U2AF homology motif; see below).

An RRM is a nucleic acid binding domain found in all organisms and typically binds single-stranded RNA, usually by recognition of a specific nucleotide sequence. RRMs are a subclass of the ferredoxin fold [16] that contain a four-stranded β-sheet buttressed by two α-helices in a β-α-β-β-α-β arrangement; they are distinguished by the presence of two amino acid motifs, RNP1 and RNP2, featuring conserved aromatic residues. High resolution structures of RRM domains both alone and bound to RNA have been described [17–21]. These reveal that single-stranded RNA typically binds on the face of the β-sheet, and that the two RNP motifs within the β-sheet are important for this interaction [22]. In particular, a tyrosine or phenylalanine within the RNP motif makes a stacking interaction with bound nucleotidyl bases while the identity of the nucleotide is determined by hydrogen bonding interactions. Structural analyses of RNA binding by the polypyrimidine tract binding protein (PTB) and the alternative splicing factor Fox-1 reveal further complexity in RNA recognition by RRM domains [23,24]. These include the substitution of RNP aromatic·RNA contacts with a separate set of hydrophobic interactions [23] and distinct features with respect to the participation of RRM loops in RNA binding [24].