RNA Biology

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/krnb20

Regulation of circRNA biogenesis

Ling-Ling Chen\textsuperscript{ab} & Li Yang\textsuperscript{bc}

\textsuperscript{a} State Key Laboratory of Molecular Biology; Institute of Biochemistry and Cell Biology; Shanghai Institutes for Biological Sciences; Shanghai, China

\textsuperscript{b} School of Life Science and Technology; ShanghaiTech University; Shanghai, China

\textsuperscript{c} Key Laboratory of Computational Biology; CAS-MPG Partner Institute for Computational Biology; CAS Center for Excellence in Brain Science; Shanghai Institutes for Biological Sciences; Chinese Academy of Sciences; Shanghai, China

Accepted author version posted online: 06 Mar 2015.

To cite this article: Ling-Ling Chen & Li Yang (2015) Regulation of circRNA biogenesis, RNA Biology, 12:4, 381-388

To link to this article: http://dx.doi.org/10.1080/15476286.2015.1020271

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at http://www.tandfonline.com/page/terms-and-conditions
Regulation of circRNA biogenesis

Ling-Ling Chen1,2,* and Li Yang2,3,*
1State Key Laboratory of Molecular Biology; Institute of Biochemistry and Cell Biology; Shanghai Institutes for Biological Sciences; Chinese Academy of Sciences; Shanghai, China; 2School of Life Science and Technology; ShanghaiTech University; Shanghai, China; 3Key Laboratory of Computational Biology; CAS-MPG Partner Institute for Computational Biology; CAS Center for Excellence in Brain Science; Shanghai Institutes for Biological Sciences; Chinese Academy of Sciences; Shanghai, China

Unlike linear RNAs terminated with 5′ caps and 3′ tails, circular RNAs are characterized by covalently closed loop structures with neither 5′ to 3′ polarity nor polyadenylated tail. This intrinsic characteristic has led to the general under-estimation of the existence of circular RNAs in previous polyadenylated transcriptome analyses. With the advent of specific biochemical and computational approaches, a large number of circular RNAs from back-spliced exons (circRNAs) have been identified in various cell lines and across different species. Recent studies have uncovered that back-splicing requires canonical spliceosomal machinery and can be facilitated by both complementary sequences and specific protein factors. In this review, we highlight our current understanding of the regulation of circRNA biogenesis, including both the competition between splicing and back-splicing and the previously under-appreciated alternative circularization.

Covalently closed circular RNAs were originally identified in plant viroids,1 yeast mitochondrial RNAs,2 and hepatitis δ virus.3 Later, circular RNAs were sporadically uncovered from either exons4-6 or introns,7,8 which were thought as by-products of spliceosome-mediated splicing errors (mis-splic- ing with scrambled exon orders)9,10 or intermediates escaped from intron lariat debranching,7,8 thus they are unlikely to play important roles in biological processes.

Being covalently closed, without polyadenylation at their 3′ ends,11,12 most circular RNAs have fallen below the radar of transcriptomic polyadenylated RNA profiling.13,14 Recently, with the depletion of both polyadenylated RNAs and rRNAs, non-polyadenylated human RNAs could be enriched for RNA sequencing (RNA-seq) analysis and unexpected RNA-seq signals were accumulated in certain exons (called excised exons) or introns (called excised introns).15 These non-polyadenylated signals were further confirmed as circular RNAs from back-spliced exons (circRNAs)16 or stabilized introns, having escaped debranching (circular intronic RNAs, ciRNAs).17 Nowadays, circRNAs have been extensively identified among a variety of transcriptomes.12,18,23 Although most of them are lowly expressed, some circRNAs have been proven to be more abundant than their linear counterparts.20 So far, endogenous circRNAs are not reported to associate with ribosomes for translation,18,21 indicating that they have a tendency to function as a new class of long noncoding RNAs (lncRNAs). However, exogenous circRNAs engineered with internal ribosome entry site (IRES) elements can be translated in vitro24 or in vivo25, thus we can not exclude the possibility that some circRNAs are translatable.

The biogenesis of circRNAs via back-splicing is different from the canonical splicing of linear RNAs.12 Interestingly, it is also distinct from the formation of other types of circular RNAs, including but not limited to those generated by direct single strand RNA ligation,26 derived from intermediates of processed rRNAs27 or circularized introns escaped from debranching7,8,17 (reviewed by Lasda and Parker28). Here, we highlight recent advances in our understanding of circRNA biogenesis, with a particular focus on its regulations and the competition between back-splicing for circular RNAs and canonical splicing for linear ones.

Keywords: alternative circularization, back-splicing, circularization, circular RNA, complementary sequence, splicing

*Correspondence to: Ling-Ling Chen; Email: linglingchen@sibcb.ac.cn; Li Yang; Email: liyang@picb.ac.cn
Submitted: 01/29/2015
Revised: 02/11/2015
Accepted: 02/13/2015
http://dx.doi.org/10.1080/15476286.2015.1020271

www.tandfonline.com RNA Biology 12:4, 381--388; April 2015; © 2015 Taylor & Francis Group, LLC

Downloaded by [Institute of Plant Physiology and Ecology] at 22:30 09 April 2015
Canonical Splicing and Back-Splicing

Canonical eukaryotic pre-mRNA splicing is catalyzed by the spliceosomal machinery to remove introns and join exons. Together with other co-/post-transcriptional processing, such as 5' capping and 3' polyadenylation, canonical splicing leads to the formation of a linear RNA transcript with 5' to 3' polarity (Top, Fig. 1). Different from canonical splicing that joins an upstream (5') splice donor site with a downstream (3') splice acceptor site, back-splicing ligates a downstream splice donor site reversely with an upstream splice acceptor site, resulting in a covalently closed circRNA transcript and an alternatively spliced linear RNA with skipped exons (Bottom, Fig. 1). However, evidence for co-expression of endogenous circRNAs and putative linear RNAs with exon exclusion is absent in most circRNAs, possibly due to their fast degradation. Nevertheless, it has been shown that both canonical splice signals and canonical spliceosomal machinery are required for back-spliced circularization. In addition, the majority of highly expressed circRNAs are usually processed from internal exons of pre-mRNAs and normally contain multiple exons, indicating that back-splicing is generally coupled with canonical splicing.

The coupling between splicing and back-splicing can be exemplified by a simple case illustrated in Fig. 1. A canonical alternative splicing event (canonical splicing 1) joins 2 non-sequential exons together to form a linear RNA with skipped middle exons. This is coupled with the covalent circularization of the skipped middle exons by back-splicing (back-splicing 2). Very often, these 2 skipped middle exons are further processed by another canonical splicing event (canonical splicing 3) to remove the intron between them (discussed below). Two models have been proposed to explain the underlying mechanism of back-spliced circularization.

Broadly speaking, the notable difference between these 2 models is which step, canonical splicing or back-splicing, happens first (Fig. 2).

Which Comes First, Canonical Splicing or Back-Splicing

If canonical splicing happens first, it will initially generate a linear RNA with skipped exons and a long intron lariat containing these skipped exons, which are then back-spliced to generate a circRNA (Fig. 2A). Thus, this model is referred to as “exon skipping” or “lariat intermediate” model. On the other hand, if the back-splicing happens first, it will directly generate a circRNA together with an exon-intron(s)-exon intermediate, which can be further processed to produce a linear RNA with skipped exons or be potentially degraded (Fig. 2B). Accordingly, the second model has been highlighted as “direct back-splicing” model.

In both models, the expression of a circRNA is theoretically associated with an alternatively spliced linear RNA with exon exclusion (Fig. 2). However, although it has been exemplified in recapitulated assays, the correlation of endogenous circRNAs with alternatively spliced linear RNAs could only be identified in some, but not all, circRNA regions. Thus, to what extent the correlation of circularization with exon skipped splicing remains to be further investigated in endogenous circumstance. Furthermore, it is still unclear under which condition(s) the spliceosomal machinery chooses either canonical splicing (Fig. 2A) or back-splicing (Fig. 2B) to start with the production of a circRNA. These two steps may happen either stochastically or even synergistically.

Cis-Elements and Trans-Factors Regulate the Competition Between Canonical Splicing and Back-Splicing

Albeit more stable as covalently closed circles and catalyzed by canonical spliceosomal machinery, most circRNAs are less abundant than their linear mRNA counterparts. One possible explanation is that back-splicing is unfavorable for spliceosome assembly and thus less efficiently catalyzed by the spliceosomal machinery. To overcome this natural disadvantage, both cis-elements and trans-factors are required to bring the downstream donor and upstream acceptor sites close together to promote back-splicing. For examples, several recent studies have reported that the processing of circRNAs can be facilitated by either RNA pairing of reversely complementary sequences across their flanking introns or within a single individual intron to promote canonical splicing for linear RNA formation (Fig. 3A). Alternatively, protein factors binding to pre-mRNAs to bridge flanking introns together (Fig. 3B). RNA pairing across flanking introns is an efficient way to promote exon circularization. RNA pairing can be formed by either repetitive elements, such as the very abundant Alu elements in human, or non-repetitive but complementary sequences. Although short sequences (as small as 30 to 40 nucleotides) were observed to be able to sufficiently facilitate circRNA biogenesis, a strong pairing capacity could dramatically enhance the production of circRNAs.

However, it is worth noting that RNA pairing is not always formed across flanking introns to facilitate back-splicing for circRNAs (top, Fig. 3C), but can also be formed within an individual intron to promote canonical splicing for linear RNA formation (bottom, Fig. 3C). For example, a conserved RNA pairing within a long intron was reported to bridge the upstream splice donor site close to the downstream splice acceptor site, and thus could promote a canonical splicing for Rbfox-mediated exon inclusion. Thus, the selection of RNA pairing across flanking introns or within a single individual intron leads to competition between back-splicing for circRNAs and canonical splicing for linear RNAs, which has been predicted to occur genome-wide and demonstrated in recapitulated assays. It should be noted that being under regulatory control, the competition can be very dynamic, leading to different expression patterns of circRNAs and linear RNAs.

So far, 2 RNA binding proteins, muscleblind (MBNL1) and Adenosine deaminase 1 acting on RNA (ADAR1), have been reported to play a role in
circRNA biogenesis, doing so by different modes of action. MBNL1 was shown to bind to its own pre-mRNA and bridge 2 flanking introns close together (Fig. 3B) to induce back-splicing, resulting in up-regulated circRNA formation from its own RNA. ADAR1 knockdown specifically up-regulates some circRNA expression, implying a role of ADAR1 in suppression of circRNA biogenesis. Mechanistically, it was suggested that such regulation is associated with Adenosine-to-Inosine (A-to-I) editing. Double strand RNA (dsRNA) pairing structures are known as A-to-I RNA editing targets by ADARs. In the normal condition, highly enriched A-to-I editing in dsRNA regions could diminish RNA pairing.
structures, resulting in reduced RNA pairing and thus less efficient backsplicing for circRNA formation. Whereas, with the reduced level of A-to-I editing after ADAR1 knockdown, RNA pairing across flanking introns is more stable and favors backsplicing for circRNA production. Notably, ADAR1 was recently shown to act as a dsRNA binding protein to interfere with microRNA processing, thus it is also possible that ADAR1 may regulate circRNA formation directly through its dsRNA binding activity, independent of RNA editing. As there are hundreds (and maybe thousands) of RNA binding proteins, it will be of interest to identify additional trans-factors involved in circRNA formation.

Figure 2. Two possible models for circRNA formation. (A) The “exon skipping” or “lariat intermediate” model for circRNA formation. The processing starts with canonical splicing for a linear RNA with skipped exons and a long intron lariat containing these skipped exons (blue bars), which is then further back-spliced to form a circRNA. (B) The “direct back-splicing” model for circRNA formation. The processing starts with back-splicing for a circRNA together with an exon-intron(s)-exon intermediate, which can be further processed to produce a linear RNA with skipped exons or to be potentially degraded. ss, splice site. BP, branchpoint.
discovery that multiple circular RNAs can be identified from a single gene locus, a phenomenon referred to as alternative circularization. There are several ways to generate different alternative circularization events. First of all, multiple circRNAs can be processed from a single gene locus with different numbers of exons included, possibly due to the competition of RNA pairing across different intron sets. For instance, the existence of multiple human circRNAs from a single gene locus is associated with a variety of Alu pairing in human introns (red arcs, Fig. 4A), suggesting a role of competition between Alu pairing in the formation of alternative circularization.

In fact, the competition of RNA pairing across different flanking introns can be very complicated, due to the fact that the competition of RNA pairing happens either within repetitive or non-repetitive elements, and that the RNA pairing can also be regulated by additional factors, such as RNA binding proteins. In addition, multiple circRNAs can be produced from a single gene locus with the internal intron included or excluded (Fig. 4B). Interestingly, the level of internal intron retention within a circRNA can be very different when

**Figure 3.** The competition of RNA pairing for splicing or back-splicing. (A) (B) Both cis-elements (A) and trans-factors (B) can affect back-splicing efficiency by taking the downstream splice donor and upstream acceptor sites close together. (C) The competition model of RNA pairing. Top, RNA pairing formed across flanking introns promotes back-splicing, leading to the formation of a circRNA and a linear RNA with exon exclusion. Bottom, the RNA pairing formed within one individual intron promotes the canonical splicing, resulting in a linear RNA with exon inclusion, but no back-splicing. Red arrows, complementary sequences. ss, splice site. BP, branchpoint.
examined from various cell lines, suggesting that the inclusion of internal introns is under regulatory control.20 Finally, multiple circular RNAs can also be generated from either exons or introns in a single gene locus (Fig. 4C, and data not shown) by distinct circular RNA formation pathways. For instance, a circular intronic RNA is produced from intron lariats that escape debranching (pink circle, Fig. 4C) and the production of such circular intronic RNAs is highly associated with consensus motifs near to 5' splice sites and branchpoint sites (pink bars, Fig. 4C).17 While in the same locus, an additional circular RNA from back-spliced exons can be simultaneously generated (blue circle, Fig. 4C), resulting in multiple circular RNAs from a single gene locus.

Although endogenous conditions and regulations for alternative circularization are complex and require further investigation, the identification of alternative circularization definitely expands our understanding of backsplicing and its regulation.

**Future Remarks**

Genome-wide annotations have revealed that circRNAs are widely expressed in different cell lines and across various species,12,18,19,22,23,31 greatly expanding our understanding of the ever growing list of IncRNAs. Recent studies have uncovered that the processing of circRNAs is not only catalyzed by canonical spliceosomal machinery,23,30 but also regulated as canonical splicing by both cis-elements12,18,23,31 and trans-factors.23,29 Given the discrete expression of circRNAs, additional cis-/trans-regulators involved in circRNA biogenesis are required to be further defined.

A challenging unanswered question is the function of circRNAs. Recently, a handful of circRNAs have been shown to be involved in gene expression regulation as miRNA sponges.19,21,40,41 Strikingly, some circular intronic RNAs and circRNAs with retained introns were also reported to function as positive regulators of RNA Pol II transcription in the nucleus.17,39 In addition, it has been proposed that some circRNAs may “sponge” other factors, such as RNA binding proteins.42 The
processing of circRNA biogenesis itself might also regulate the formation of linear mRNAs through the competition between splicing and back-splicing. Since dozens of thousands of circRNAs have been identified, in the future, it will be of great interest to study what the vast majority of circRNAs can do in both physiological and pathological conditions.

Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed.

Acknowledgments
We are grateful to Andrew E. Teshchenko for critical reading of this manuscript. We apologize to authors whose work could not cite here owing to space/permission limitations.

Funding
Our work is supported by 31471241, 31322018 and 91440202 from NSFC.

References
1. Sanger HL, Klitz G, Riessner D, Gross HJ, Kleinschmidt AK. Virions are single-stranded covalently closed circular RNA molecules existing as highly base-paired rod-like structures. Proc Natl Acad Sci U S A 1976; 73: 3852-6; PMID:10509269; http://dx.doi.org/10.1073/pnas.73.11.3852
3. Kos A, Djikema R, Arnberg AC, van der Meide PH, Schellekens H. The hepatitis delta (delta) virus possesses a circular RNA. Nature 1986; 325:558-60; PMID:2429129; http://dx.doi.org/10.1038/325558a0
7. Kopczynski CC, Musikavich MA. Introns excised from the Delta primary transcript are localized near sites of Delta transcription. J Cell Biol 1992; 119: 503-12; PMID:1383233; http://dx.doi.org/10.1083/jcb.119.3.503


42. Hentze MW, Preiss T. Circular RNAs: splicing’s enigma variations. EMBO J 2013; 32: 923–5; PMID:23463100; http://dx.doi.org/10.1038/emboj.2013.53