Available online at www.sciencedirect.com
SciVerse ScienceDirect
Journal of Genetics and Genomics 39 (2012) 535−543
JGG

REVIEW

The Role of RNA Structure in Posttranscriptional Regulation of Gene Expression

Elina Jacobs[1], James D. Mills, Michael Janitz*

School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney NSW 2052, Australia

Received 27 June 2012; revised 16 August 2012; accepted 17 August 2012
Available online 5 September 2012

ABSTRACT

As more information is gathered on the mechanisms of transcription and translation, it is becoming apparent that these processes are highly regulated. The formation of mRNA secondary and tertiary structures is one such regulatory process that until recently it has not been analysed in depth. Formation of these mRNA structures has the potential to enhance and inhibit alternative splicing of transcripts, and regulate rates and amount of translation. As this regulatory mechanism potentially impacts at both the transcriptional and translational level, while also potentially utilising the vast array of non-coding RNAs, it warrants further investigation. Currently, a variety of high-throughput sequencing techniques including parallel analysis of RNA structure (PARS), fragmentation sequencing (FragSeq) and selective 2-hydroxyl acylation analysed by primer extension (SHAPE) lead the way in the genome-wide identification and analysis of mRNA structure formation. These new sequencing techniques highlight the diversity and complexity of the transcriptome, and demonstrate another regulatory mechanism that could become a target for new therapeutic approaches.

KEYWORDS: Transcriptome; mRNA structure; Parallel analysis of RNA structure; Fragmentation sequencing; PARS; FragSeq; SHAPE-Seq

1. INTRODUCTION

RNAs are an integral part of the cellular composition and involved in many important processes. While RNA was first thought to mainly be the mediator between DNA and proteins, with the exceptions of ribosomal RNA (rRNA) and transfer RNA (tRNA), it is now known that RNA molecules can have many more functions (Mattick and Makunin, 2006). Considering the many steps involved in RNA production, transcription, splicing, transportation and translation, there is a lot of room for regulation of RNA molecules. It is becoming increasingly apparent that the information for this regulation is not confined to the primary sequence of the messenger RNA (mRNA), but may also be given through higher-order structures (Chen et al., 1999). This allows for an increased amount of information to be stored within DNA without increasing genome size (Huynen et al., 1993). Local nucleotides can form secondary structures such as hairpins and stem-loops through base-pairing, and even more distantly located sequences can interact and form tertiary structures (Wan et al., 2011). Assessing RNA structure is an important part of determining the functional assignment of a gene product (Floris et al., 2011). However, our current knowledge about the structural transcriptome is still very limited, largely due to the low-throughput nature of RNA structure probing (Wan et al., 2011). Nevertheless, it is now appreciated that secondary and tertiary structures can have an impact on all classes of RNAs and in all aspects of their functions (Wan et al., 2011). This short review will focus on what is currently known about

Abbreviations: SMN, survival motor neuron; MAPT, microtubule-associated protein tau; m7G, 7-methyl-guanylate; CAT, chloramphenicol acetyltransferase; PARS, parallel analysis of RNA structure; FragSeq, fragmentation sequencing; SHAPE, selective 2-hydroxyl acylation analysed by primer extension.

* Corresponding author. Tel: +61 2 9385 8608, fax: +61 2 9385 1483.
E-mail address: m.janitz@unsw.edu.au (M. Janitz).
[1] Present address: College of Science and Engineering, University of Edinburgh, Edinburgh EH9 3JR, United Kingdom.

http://dx.doi.org/10.1016/j.jgg.2012.08.002

the role of mRNA structures in regulation of gene expression, with a focus on higher eukaryotes and humans. In particular the review will focus on alternative splicing (AS) since this control mechanism of genetic information flow is currently considered to be the most versatile in its potential to increase proteome complexity from relatively limited number of protein-coding genes in higher vertebrates. Furthermore, the latest tools for measuring and mapping RNA secondary structures will be briefly discussed, and indications for future developments will be given.

2. ALTERNATIVE SPLICING

Alternative splicing allows the generation of multiple transcripts from one gene, adding another source of diversity to the transcriptome and proteome (Black, 2003). Since the discovery of AS in the adenovirus it was soon realised that this mechanism is also extensively utilised by eukaryotes, including mammals and humans (Berget et al., 1977; Chow et al., 1977; Hallegger et al., 2010). The increase in transcriptional data has been accompanied by an increase in known alternative transcripts, and it has been estimated that up to 95% of all human genes are alternatively spliced (Pan et al., 2008). Much research has been done to elucidate the role of this process, however, the initial idea that all alternative transcripts are functionally important has been challenged. While there are well-established examples of the functional relevance of alternatively spliced products, evidences that the majority of alternative transcripts are expressed at low levels, with many leading to unstable protein conformations, are emerging (Modrek and Lee, 2003; Pan et al., 2004; Melamud and Moult, 2009b; Hegyi et al., 2011). Furthermore, most protein isoforms have conserved structure and function, suggesting that there is selection against structurally different isoforms, which in turn suggests selection against alternative transcripts leading to structurally different isoforms (Ezkurdia et al., 2012). Indeed, it has been suggested that noise in the splicing process may be responsible for most alternative transcripts (Melamud and Moult, 2009a). Yet, it has become increasingly apparent that RNAs themselves perform many important functions without being translated into proteins (Mattick and Makunin, 2006). Many of these non-coding RNAs originate from alternatively spliced transcripts that can be structurally very different from each other (Floris et al., 2011). While it cannot be excluded that some isoforms are the result of splicing noise and are non-functional, there is a possibility that the differences in structure lead to unrelated functions. For example it has been found that various transcripts of the phosphatase and tensin homolog (*PTEN*), a tumour suppressor gene, are biologically active, competing for the same microRNA (miRNA) binding site (Poliseno et al., 2010). This further highlights the importance of considering alternative mechanisms when looking at mRNA function.

The splicing process is carried out by the spliceosome, a nuclear macromolecule complex that removes introns from pre-mRNAs in conjunction with additional auxiliary factors (Warf and Berglund, 2010). The spliceosome recognises specific sequences at the 5′ and 3′ splice sites of the intron, a branch-point contained within the intron and a polypyrimidine tract, and then cleaves the intron accordingly. While these sequences tend to be highly conserved in less complex organisms, they are less conserved in higher eukaryotes, suggesting that sequence information may not be the sole element determining the regulation of splicing. Furthermore, higher eukaryotes have additional splicing regulatory elements (SREs), such as exonic splicing enhancers (ESEs) or exonic splicing silencers (ESSs), and intronic splicing enhancers (ISEs) or intronic splicing silencers (ISSs) (Wang and Burge, 2008). It is so far thought that the regulation of AS in higher eukaryotes occurs *via* different sets of RNA binding proteins that recognise splice sites and branching points as well as different enhancer or silencer elements within or surrounding alternatively spliced exons (Hallegger et al., 2010). However, the fundamental rules of AS are yet to be established.

While most research has focused on identifying the sequences that may mediate the splicing information, it is becoming increasingly apparent that pre-mRNA structures can play an important role. This may be mediated by several mechanisms: bringing distant splice signals into closer proximity, or affecting accessibility of splice sites by either masking or exposing them (Singh et al., 2007).

2.1. mRNA structures that alter splicing patterns

2.1.1. Human growth hormone (*GH1*)

There are two isoforms of the *GH1* that are differentially expressed in the pituitary and the placenta. In the pituitary, both isoforms are expressed, whereas only the major transcript is expressed in the placenta. This difference in expression is thought to be regulated by mRNA secondary structures. A predicted stem-loop formation that encompasses the minor splice site and its lariat branch-point are associated with increased expression of the minor isoform, thus suggesting that the formation of this secondary structure may allow for expression of this isoform. Since the minor isoform is expressed at lower levels, it has been suggested that its secondary structure exists in equilibrium with other structures that favour the use of the splice site of the major isoform (Estes et al., 1992). Mutagenesis experiments that alter the intronic sequence but not the secondary structure of the major isoform confirm this suggestion. The secondary structures are thought to affect the accessibility of the alternative splice sites (Estes et al., 1992). Although the predicted structures remain to be confirmed experimentally and it is unclear why they are tissue specific, this provides an example of how secondary structures may regulate the accessibility of splice sites and thereby regulate the expression of alternative transcripts.

2.1.2. Survival motor neuron (*SMN*)

There are two *SMN* isoforms, *SMN1* and *SMN2*, which differ in their inclusion of exon 7. The difference lies in a C to T substitution in exon 7 that does not affect the coding sequence but changes the splicing pattern by introducing an inhibitory element to the 3′ splice site in *SMN2*, which causes