Biochimie 93 (2011) 1922–1927

Contents lists available at ScienceDirect

Biochimie

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

Review

Non-coding RNA and antisense RNA. Nature's trash or treasure?

Stuart Knowling*, Kevin V. Morris

Department of Molecular and Experimental Medicine, The Scripps Research Institute, 10550 N. Torrey Pines Road, La Jolla, CA 92037, USA

ARTICLE INFO

Article history:
Received 31 May 2011
Accepted 29 July 2011
Available online 10 August 2011

Keywords:
Antisense RNA
Non-coding RNA
Epigenetic
Transcription
Gene silencing

ABSTRACT

Although control of cellular function has classically been considered the responsibility of proteins, research over the last decade has elucidated many roles for RNA in regulation of not only the proteins that control cellular functions but also for the cellular functions themselves. In parallel to this advancement in knowledge about the regulatory roles of RNA there has been an explosion of knowledge about the role that epigenetics plays in controlling not only long-term cellular fate but also the short-term regulatory control of genes. Of particular interest is the crossover between these two worlds, a world where RNA can act out its part and subsequently elicit chromatin modifications that alter cellular function. Two main categories of RNA are examined here, non-coding RNA and antisense RNA both of which perform vital functions in controlling numerous genes, proteins and RNA itself. As the activities of non-coding and antisense RNA in both normal and aberrant cellular function are elucidated, so does the number of possible targets for pharmacopeic intervention.

1. RNA, regulatory nucleic acid?

Recent research using techniques such as Genomic tiling arrays and cDNA sequencing has allowed transcription in the human genome to be studied with a degree of accuracy and resolution previously unattainable. The ENCODE project identified several functional elements in the Human genome and demonstrated that RNA can be processed to yield both short and long RNAs, which can overlap the 5′ and 3′ ends of protein-coding regions [1]. Although approximately 90% of the Human genome is transcribed [2] the ENCODE project demonstrated that a surprisingly small amount actually encodes protein (~2% of the eukaryotic genome encodes protein-coding genes (mRNA)), therefore, a vast number of transcripts appear which are non-protein-coding RNAs (ncRNAs).

The identification of these long or short ncRNAs effectively split DNA transcription into two camps, coding and non-coding transcripts [3] where traditional non-coding transcripts are viewed to be ribosomal RNA (rRNA) and transfer RNA (tRNA), both of which play an essential function in protein translation. As for the other ncRNAs, which can be either long (>200 bp) or short, there is disagreement as to whether these ncRNAs are transcriptional "noise": nevertheless there is growing evidence that ncRNAs play an important role in cellular function [3] as is the case for the aforementioned tRNA and rRNA.

In general, the more complex an organism, the greater its number of ncRNAs [4]. Supporting this varied complexity is the observation that the processing of longer RNA to shorter RNA to yield unique secondary and tertiary structures that participate in cellular processes has long been known to occur in the presence of [5] and absence of protein [6,7]. The enticing possibility that although the level of protein-coding transcripts between organisms is similar, the ultimate control of cellular function may be through interactions between proteins and ncRNA, which is corroborated by the fact that the majority of chromatin-modifying complexes do not have DNA binding capacity [8] and therefore, must utilize a third party in binding to DNA. This is in stark contrast to transcription factors, which specifically recognize and bind DNA sequences and do not require extrinsic factors.

Classically it has been thought that control of gene expression is largely due to the action of transcription factors but ncRNAs have also recently been shown to exert their control over gene transcription *via* several different pathways, transcriptional gene silencing (TGS) through the targeted recruitment of epigenetic silencing complexes to particular loci (reviewed in [9,10]), through post-transcriptional gene silencing (PTGS); degradation of transcriptionally active mRNAs as exhibited in RNAi, siRNA and miRNA, and also via STAU-1 mediated RNA decay process [11]. Control of gene expression by ncRNA is evolutionary sensible; the response is rapid due to the speed of ncRNA production in the vicinity of the gene and the energy costs to the cell are much lower due to the lack of protein synthesis.

* Corresponding author. Tel.: +1 858 784 2812; fax: +1 858 784 8836.
E-mail address: knowling@scripps.edu (S. Knowling).

0300-9084/$ – see front matter © 2011 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.biochi.2011.07.031

2. Epigenetics and ncRNAs

Epigenetics is the study of the underlying changes in phenotype that are caused by alterations to the expression of the genome by chemical modification of the DNA molecule. Notably, these chemical changes to the DNA do not alter the sequence context of the DNA. Although several forms of epigenetic regulation exist, the two main forms of interest for this review are the (1) addition of chemical groups to specific bases, as with DNA methylation and (2) the local alterations of histones, such as targeted methylation at particular lysines that affects the accessibility of the surrounding genomic DNA to the transcriptional machinery. These epigenetic controls have been shown to be capable of being passed onto daughter cells as shown by changes to the *IGF2* gene that are persistent across familial generations six decades later [12]; an effect that has also been noted in mice [13,14]. One obvious question, which arises from such observations, is whether or not there is a link between the large amount of transcribed ncRNA and the regulation of genome modification *via* epigenetics. Given that high-throughput sequencing revealed transcription in ~90% of the genome, including RNAs transcribed at a low levels and not rapidly degraded [2]; suggests there must be a function for these ncRNAs, *i.e*. this is an energy cost to the cell and these transcripts have been retained over successive generations, so one would expect that they play some, as yet unknown, beneficial role. Although we currently cannot exclude the fact that the mere act of transcription of ncRNA is an archaic remnant of cellular activity and is due to the presence of a promoter region; growing evidence suggests that at least some of these ncRNA molecules play specific roles in eukaryotic cells and gene expression (reviewed in [9]). The comprehensive role that ncRNA plays in this epigenetic control remains to be fully elucidated but if recent observations are any indication [15–22], ncRNAs in human cells might be active regulators involved in controlling gene expression *via* the targeted recruitment of epigenetic complexes to various loci in the genome.

Expressed ncRNAs can show clear evolutionary conservation [23] and many emanate from gene promoter regions, which tend to be more conserved than protein-coding genes [16] but also exhibit little conservation of expressed regions between different species [16,24,25]. These observations suggest a level of retention in the machinery of the cell and a possibly harkening to a role in gene level control. It should be noted that although ncRNAs can be found in polyadenylated, unadenylated or bimorphic forms [26] and do not contain classical ORFs longer than 100 amino acids; some might in fact encode small peptides [27]. Such an eventuality could add even more layers of complexity to the cell than have been previously appreciated.

3. Antisense RNA and non-coding RNA: the Yin and Yang of gene control

Some of the most studied ncRNA to date have been the long intergenic non-coding RNAs (lincRNAs), which are a heterogeneous group of transcripts involved in epigenetic control of the cell that range in size from ~300 nucleotides to several thousands. Currently the human catalog of lincRNAs is thought to be around 3300 although the true number may be closer to 4500 [28]. Often associated with these ncRNA is an antisense RNA (asRNA) that contains a sequence complementary to the ncRNA and thus may afford the cell another layer of genetic regulation.

To date the most studied and well understood lincRNA is the 17,000 nucleotide transcript Xist, which is involved in X chromosome inactivation (for an in-depth review see [29]). Of prime importance in X-inactivation is the X-inactivation center (XIC in humans, Xic in mice), which contains at least two ncRNA, the aforementioned XIST (XIST humans and Xist in mice) and its asRNA Tsix. Expressed early on in embryonic development, Xist is weakly expressed by both X chromosomes until cell differentiation when an yet-to-be determined key factor triggers up-regulation of Xist transcription from the future inactive chromosome by progressive coating of the chromosome from the XIC outwards [30]. In humans this randomly coats one of the two X chromosomes in females whereas in mice the Xist locus on the maternal X chromosome is always repressed and thus, the maternal X chromosome is always active giving rise to an X_{active} and $X_{inactive}$ [31]. Upon differentiation the histone modification of the active and inactive become significantly altered with the inactive X chromosome exhibiting more repressive chromatin modification, which is thought to play a role in recruitment of proteins, while the active X chromosome exhibits silencing of the Tsix asRNA promoter due to a lack transcriptional machinery recruitment. The result of these eventualities is the alteration of the expression of Xist and the coating of one of the chromosomes by the ncRNA causing inactivation of those chromosome associated genes due to the loss of histone modification by acetylation and methylation [32]. This coating of the chromosome ensures an equal dosage of gene expression between X-linked genes of males and females. In mice, this inactivation has been shown to require an interaction between the 5′ of Xist, named RepA [33] and the Polycomb Repressive Complex 2 (PRC2, a complex containing histone methyltransferases (HMTases), Enhancer of Zeste (EZH2, a H3K27 histone methyltransferase) and SUZ12 or G9A (both of which are H3K9 histone methyltransferases)).

Therefore, in order to maintain an active X chromosome, control of the Xist ncRNA is critical. In mice this is achieved by the action of the antisense RNA Tsix. Notably, in Humans TsiX appears to either truncated or absent [34]. Tsix is a 40 Kb antisense transcript (that is further processed to smaller transcripts [35]) and found to be expressed from the X controlling element at the opposite strand of Xist and thus overlaps the 3′ end of Xist gene [36]. Unlike Xist, Tsix shows high levels of transcriptional machinery and epigenetic marks in undifferentiated cells [37]. Tsix acts in *cis* to regulate the role of Xist in X-inactivation by repressive marking of the Xist promoter. The repressed state of Xist exhibits low levels of (H3K4) methylation and (H3K9) acetylation whilst also exhibiting higher levels of the repressive H3K9 and CpG DNA methylation [37]. Due to these epigenetic marks the Xist RNA cannot be detected by either ChIP or nuclear run on transcription [38,39] and ensures that the chromosome is not coated in Xist; although doubt still surrounds this hypothesis [40] it appears most likely that regulation of Xist is due to a shift from a heterochromatin state to a euchromatin state. Interestingly, the human TSIX contains only ~50% sequence homology to its murine cousin, but retains many similarities in that it is expressed in embryonic derived cells, produces a ncRNA and is initiated downstream 3′ of XIST. As the human TSIX transcript does not appear to overlap the 5′ region of XIST, such an eventuality rules out a direct interaction with the XIST promoter. What role human TSIX exactly plays remains to be determined. Some have postulated that the human TSIX is simply an evolutionary carry over [34].

Although not as well studied as X-inactivation, one area of epigenetics that utilizes ncRNA that are antisense to their gene counterpart is gene imprinting, where genes are expressed in a manner that is dependant upon their paternal or maternal origin. Some of the first genes to be identified as imprinted were the insulin-like growth factor type-2 receptor (Igf2r), Igf2 and H19 locus [41–43]. H19 is a tumor-suppressor gene and produces a long ncRNA (although considerably shorter than Xist) that is 2.3 kb long, capped and polyadenylated [44]. The H19 ncRNA is transcribed exclusively from the maternal allele during development of the embryo but down-regulated after birth, whilst the IGF2 gene is expressed exclusively from the paternal allele [43].