The non-coding RNAs as riboregulators

Volker A. Erdmann, Mirosława Z. Barciszewska¹, Maciej Szymanski¹, Abraham Hochberg², Nathan de Groot² and Jan Barciszewski¹,*

Institute of Biochemistry, Freie Universität Berlin, Thielallee 63, 14195 Berlin, Germany, ¹Institute of Bioorganic Chemistry of the Polish Academy of Sciences, Noskowskiego 12, 61704 Poznan, Poland and ²Department of Biological Chemistry, Institute of Life Sciences, The Hebrew University, IL-91904 Jerusalem, Israel

Received October 10, 2000; Revised and Accepted October 17, 2000

ABSTRACT

The non-coding RNAs database (http://biobases.ibch.poznan.pl/ncRNA/) contains currently available data on RNAs, which do not have long open reading frames and act as riboregulators. Non-coding RNAs are involved in the specific recognition of cellular nucleic acid targets through complementary base pairing to control cell growth and differentiation. Some of them are connected with several well known developmental and neurobehavioral disorders. We have divided them into four groups. This paper is a short introduction to the database and presents its latest, updated edition.

INTRODUCTION

RNA plays a variety of structural, informational and regulatory roles in the cell. Some RNAs possess catalytic properties and act as ribozymes. 23S rRNA alone possesses peptidyltransferase activity and other RNAs as aptamers can bind small molecular compounds. In the last few years many data have accumulated showing that various non-translatable, non-coding RNA transcripts are present in different cells. They are lacking a protein coding capacity and it seems they exert their functions of cellular nucleic acid targets through complementary base pairing to control cell growth and differentiation. The non-coding RNAs as riboregulators

(i) Gene regulators. They affect the activity of a gene by different mechanisms such as silencing, methylation and DNA–RNA or RNA–RNA interaction. This group includes: Xist, roX, PAT-1, Tsix, XistAS, H19, IPW, NTT, DGCR5, KvLQTIAS, Nesp GNAS, SCA8, CMPD, lin-4, let-7, UBE3A, ZNF127AS, ScYc, DISC2, sok, CopA, RNAI, pnd, RNA-OUT.

(ii) Abiotic stress signals. These RNAs are synthesized in response to stress (e.g. oxidation) and include: gadd7/adapt15, adapt33, hsro, G90, OxyS, DsrA.

(iii) Biotic stress signals. This group consists of RNAs inducible by biologically active molecules (e.g. cytokinins). They are: His-1, ENOD40, lbi, CR20, GUT15.

(iv) Other RNAs within this group have a different origin and function. They include: Bsr, BC1, BC200, SRA, meiRNA, UHG, Xlsirt.

In comparison with the previous database new RNAs, especially antisense RNAs, are included (1). Their nucleotide sequences are stored and can be retrieved as separate files at: http://biobases.ibch.poznan.pl/ncRNA/. Below we characterize briefly those ncRNAs for which new data have appeared since last year’s edition of the database (1).

DOSSAGE COMPENSATION RNAs

One of the fundamental differences between male and female cells is the number of X chromosomes. The difference in its content and the requirement of equal expression of the chromosome X genes in both sexes led to the evolution of several types of dosage-compensation mechanisms. In mammals this process involves expression of specific genes, whose products (RNA transcripts) do not contain a long ORF (1–4). In Drosophila, compensation for the reduced dosage of genes located on a single male X chromosome involves doubling their expression in relation to their counterparts on female X chromosomes. This is an epigenetic process involving the specific acetylation of histone 4 at lysine 16 by histone acetyltransferase (MOF). It is part of the chromosome-associated dosage compensation complex comprising male-specific lethal (MSL) proteins and rox RNAs. The dosage compensation process involves products of the two genes roX1 and roX2, which are male specific and do not encode proteins. It is already known that roX1 RNA becomes associated with chromosome X at sites determined by binding of the msl (male specific lethal) gene products, which leads probably to the remodeling of the chromatin and allows increased transcription. The roX RNAs are expressed in all somatic cells in males and in diploid cells they co-localize with the MSL proteins (5).

A two-step process for recognition of the X chromosome by the MSL protein complex has been proposed (6). In the first step, MSL 1 and MSL 2 together recognize 30–40 sites distributed along the length of X, where they recruit other MSL proteins. Two of these sites encode roX RNAs, which are incorporated into the growing MSL complex. In the second step, the MSL complex can associate with chromatin entry sites on X

*To whom correspondence should be addressed. Tel: +48 61 852 8503; Fax: +48 61 852 0532; Email: jbarcisz@ibch.poznan.pl
and then spread along the chromosome in cis to locate all of the genes that utilize MSL-mediated dosage compensation (6–8).

### H19 RNA

In adulthood a basal H19 gene expression has been detected only in mammary gland, cardiac and skeletal muscles and to a lesser extent in kidney, adrenal gland and lung. The cells with a high H19 level stopped their proliferation after 48 h when cultivated in a low serum-containing media, while the cells lacking H19 continued their proliferation (9). It was suggested that H19 RNA could be used as an adjuvant tumor marker for the diagnosis, staging and follow-up of patients with serious ovarian carcinoma (10,11).

The imprinting is the result of methylation of a 7–9 kb domain on the paternal allele of H19. The 5’ flank of the H19 gene contains an imprinting mark region characterized by paternal allele-specific methylation. The evolutionarily conserved 42 bp element located upstream of the domain might play a role in imprinting and/or transcriptional regulation of H19 (12). The G-rich repeat 1.5 kb upstream of mouse H19 was shown to be present in rats but absent in humans, and is not essential for H19 imprinting. The transcriptionally active maternal allele is unmethylated whereas the inactive paternal allele is methylated. The differentially modified region in the mouse H19 locus extends over 2 kb at −4 kb upstream of the gene promoter. A similar pattern of methylation is observed for the human H19 gene. Its RNA contains four attachment sites for the oncofetal IGf2 mRNA-binding proteins with Kd’s of 0.4–1.3 nM. These sites are located within a 700 nt segment encoded by exons 4 and 5 (13). Recently a chromatin boundary model of genomic imprinting has been proposed. It suggests that chromatin boundary elements (insulators) act in cis at an insulated gene and its regulatory portion thereby blocks transcription when placed between a gene and its enhancer. An insulator located upstream of the H19 gene isolates Igf2 from its enhancers. They are accessible to Igf2 but not to H19 when moved between two genes (upstream of the putative insulator). When the ICR is unmethylated on the maternally inherited chromosome, there are two nuclease hypersensitive regions that overlap with several short CG rich repetitive elements conserved in human and mouse. These repeats are the target for the conserved DNA-binding zinc finger protein CTCF. The CTCF-binding factor interacts with the core insulator sequence via a domain containing 11 zinc fingers. The resulting insulator blocks activation of the maternal copy of Igf2 by H19. The methylated ICR contains no hypersensitive sites and this prevents CTCF binding, thereby inactivating the insulator and allowing the H19 enhancer to activate Igf2 (14–21). Recently it was also demonstrated that H19 gene expression is controlled by steroid hormones and this gene is highly expressed in hormone-sensitive organs when the hormonal stimulation is accompanied by morphological repair (22).

### SCA8

Spinocerebellar ataxia type 8 (SCA8) is a neurodegenerative disorder caused by the expansion of a CTG trinucleotide repeat, which is transcribed as a part of an untranslated RNA. It has been found that the SCA8 gene is transcribed through the repeat only in the CTG orientation, as in the case for myotonic dystrophy (DM), but not in the CAG orientation, as in the other dominantly inherited ataxias: SCA1, 2, 3, 6 and 7. In these latter cases, the CAG expansion is translated into a polyglutamine tract that adds a toxic gain of function to the respective proteins, whereas the CTG expansions in DM and SCA8 are not translated. The RNA transcripts containing the SCA8 CTG are alternatively spliced and polyadenylated, and are finally expressed in various brain tissues. No extended ORF has been found (27,28). The SCA8 CTG repeat is preceded by a poly-morphic but stable CTA tract with a (CTA)1–21(CTG)n configuration. The affected individuals analyzed had either an uninterrupted CTG repeat tract or an allele with one or more CCG, CTA, CTC, CCA or CTT insertions. In addition, the SCA8 repeat tract in sperm underwent contractions, with nearly all of the resulting expanded alleles having repeat lengths of <100 CTGs, a size that is not often associated with disease. These repeat contractions in sperm likely underlie the reduced penetrance associated with paternal transmission (29).

### DEVELOPMENTAL TIMING RNA

The Caenorhabditis elegans heterochronic gene pathway consists of a cascade of regulatory genes that are temporally controlled to specify the timing of developmental events. The products of the heterochronic genes include transcriptional and translational regulators and two different cases of novel small translational regulatory RNAs. Other genes of the pathway encode evolutionarily conserved proteins (30–34).


Another heterochronic switch gene is let-7. Loss of its activity causes reiteration of larval cell fates during the adult stage, whereas increased let-7 gene dosage results in precocious expression of adult fates during larval stages. let-7 encodes a temporally regulated 21 nt RNA that is complementary to elements in the 3′ untranslrated regions of lin-14, lin-28, lin-41,
lin-42 and daf-12. The sequential stage-specific expression of the lin-4 and let-7 regulatory RNAs trigger transitions in the complement of heterochronic regulatory proteins to coordinate developmental timing. lin-4 and let-7 are the only genes known to encode small RNAs that specifically regulate other genes. However, it is clear that lin-14 is regulated by lin-4 RNA at a step after transcription (35).

OTHER NON-CODING RNA TRANSCRIPTS

Non-coding antisense RNA has been found that overlaps with ubiquitin protein ligase (UBE3A) connected with Angelman syndrome, which is characterized by severe neurological features (36). It covers the 3′ half of UBE3A and additional sequence downstream. The 5′ end of the antisense transcript is 6.5 kb from the stop codon of UBE3A (37).


A balanced translocation segregating with schizophrenia can disrupt the gene function. It was found that two novel genes are disrupted by translocation: DISC1, encoding a protein and DISC2, which is a non-coding RNA lying antisense to DISC1 (40).

The myoplasm is a localized cytoplasmic region that is involved in axis determination, gastrulation, muscle cell specification and the pattern of cell division during ascidian development. One-cell zygotes of the ascidian Styela clava contain 1.2 kb polyadenylated RNA (yellow crescent RNA), present throughout embryonic development and associated with the cytoskeleton. The S.clava yellow crescent RNA (Sc Yc RNA) has no long ORF. It contains a short ORF that encodes a putative peptide of 49 amino acids without significant homology to known proteins (41). It is localized in the cytoplasm and segments to the larval muscle cells during cleavage. Probes containing the 3′ region of Yc RNA were used to identify maternal Yc-related RNAs. A cDNA clone encoding the ascidian Proliferating Cell Nuclear Antigen (PCNA) (39) contains 1.2 kb polyadenylated RNA (yellow crescent RNA), present throughout embryonic development and associated with the cytoskeleton. The PCNA mRNA (CopT). The CopA–CopT binding process is viewed as a series of reactions leading to progressively more stable complexes. CopA and CopT are fully complementary and both RNAs contain a major stem–loop structure, which is essential for high pairing rates. The initial step involves a transient loop–loop interaction (kissing complex) between complementary hairpin loops (46,47).

An antisense RNA, RNAI, interacting with the preprimer, RNAII, controls replication in the plasmids of the CoII family via initial and transient base pairing between complementary loops. From crystallographic data it turned out that the stem–loop structures have melted the duplex. In the hairpin conformation the RNA oligos bind the plasmid encoding the four-helix bundle protein rop (48).

Very similar interactions are also observed in other plasmids: pnd–pndB of R483 (144) and RNA-IN/RNA-OUT of IS10 (49).

DsR A RNA

The regulation of capsular polysaccharide synthesis in Escherichia coli K-12 depends on the level of an unstable positive regulator, RcsA. The amount of RcsA protein is small because of its rapid degradation and low level of protein synthesis. The latter effect is due to transcriptional silencing by the histone-like protein H-NS. A small, 85 nt DsRA RNA, activates transcription by counteracting H-NS silencing when overproduced. DsRA contains regions of sequence complementary to at least five different genes: hns, argR, ilvH, rpoS and rbsD. However, it acts in trans by RNA–RNA interactions with only two different mRNAs: hns and rpoS. H-NS is a major nucleoid-structuring, histone-like protein responsible for the silencing of a number of bacterial genes, and RpoS(G·) is the stationary phase and stress response sigma-factor of RNA polymerase. DsRA antagonizes H-NS function by decreasing the level of H-NS protein in the cell and increasing that of the RNA oligos.
mRNA, which is then exposed to nucleases. These interactions are proposed to circularize hns mRNA and provide a structural basis for DsrA activity at hns (50).

Ibi RNA

The Ibi (lipopolysaccharide biosynthesis interfering) RNA of phage Acm1, a non-translated RNA species of 97 nt, affects synthesis of a D-glucan component of the O-specific polysaccharide in serotype O1. Ibi RNA consists of two consecutive stem-loop structures. The 5'-proximal hairpin–loop function is a key structural element in the mechanism leading to the inhibition of D-glucan biosynthesis due to its antisense interactions with cellular target RNAs (51).

THE DATABASE

The database provides access to the published ncRNAs. The files contain the nucleotide sequences as well as other information in EMBL format. The primary structures are available from a hypertext list for each RNA. All data can be retrieved as separate files. Any suggestions are welcome.

ACKNOWLEDGEMENTS

This work was supported by the Polish Committee for Scientific Research (KBN) and by DFG Trilateral Research Project ‘Genomic Imprinting in Human Bladder Cancer’.

REFERENCES


