Review

Beyond the proteome: non-coding regulatory RNAs

Maciej Szymański and Jan Barciszewski

Address: Institute of Bioorganic Chemistry, Polish Academy of Sciences, Noskowskiego 12/14, 61-704 Poznań, Poland.

Correspondence: Jan Barciszewski. E-mail: jbarcisz@ibch.poznan.pl

Published: 15 April 2002

Genome ***Biology*** 2002, **3(5)**:reviews0005.1–0005.8

The electronic version of this article is the complete one and can be found online at http://genomebiology.com/2002/3/5/reviews/0005

Abstract

A variety of RNA molecules have been found over the last 20 years to have a remarkable range of functions beyond the well-known roles of messenger, ribosomal and transfer RNAs. Here, we present a general categorization of all non-coding RNAs and briefly discuss the ones that affect transcription, translation and protein function.

The 'central dogma of molecular biology' defined a general pathway for the expression of genetic information stored in DNA, transcribed into transient mRNAs and decoded on ribosomes with the help of adapter RNA (tRNAs) to produce proteins, which in turn perform all the enzymatic and structural functions in the cell. According to this view, RNAs play a rather accessory role and the complexity of a given organism is defined solely by the number of proteins encoded in its genome, according to the 'one gene - one protein' hypothesis. This simple picture was first complicated when the primary transcripts of eukaryotic protein-coding genes were found to have their coding sequences interrupted by introns [1], and it was realized that having introns provided a way to synthesize more than one protein product from a single gene, by alternative splicing [2].

Over twenty years ago, the discovery of the catalytic properties of the RNA subunit of ribonuclease P and the self-splicing activity of group I introns suggested that the functions of RNA go far beyond a passive role in the expression of protein-coding genes. *In vitro* selection techniques, which allow fast functional evaluation of large populations of RNA molecules, demonstrated that RNAs can be efficient catalysts ('ribozymes') [3]. Recent studies of the crystal structure of the large subunit of the bacterial ribosome, showed that ribosomal 23S RNA plays a key role in the process of peptide-bond formation during translation and demonstrated that ribosomes are in fact ribozymes [4]. All of these findings contributed to the hypothesis of a primordial 'RNA world', in which RNA molecules originally both carried information and fulfilled enzymatic functions. In the course of evolution most catalytic functions were taken over by proteins, and the major carrier of genetic information became the chemically more stable DNA.

Functional non-coding RNAs are not only molecular fossils left from a time when organisms consisted solely of RNA, however. They play important roles in modern-day organisms [5]. The analysis of sequenced genomes suggests that protein-coding genes alone are not enough to account for the complexity of higher organisms. There are fewer protein-coding genes in the eukaryotic genomes that have been completely sequenced so far than expected; the *Caenorhabditis elegans* and *Drosophila melanogaster* genomes contain only twice as many genes as yeast or some bacteria, and in the human genome the number is about twice that of invertebrates.

In proteome-oriented analyses of genomic sequences, genes that produce non-protein-coding transcripts are often ignored. From genomic analyses it is evident, however, that with increase of an organism's complexity, the protein-coding contribution of its genome decreases (Figure 1). It is estimated that about 98% of the transcriptional output of eukaryotic genomes is RNA that does not encode protein [6]; this includes introns and transcripts from non-protein-coding genes, with the latter accounting for 50-75% of all transcription in higher eukaryotes [7,8]. In addition to

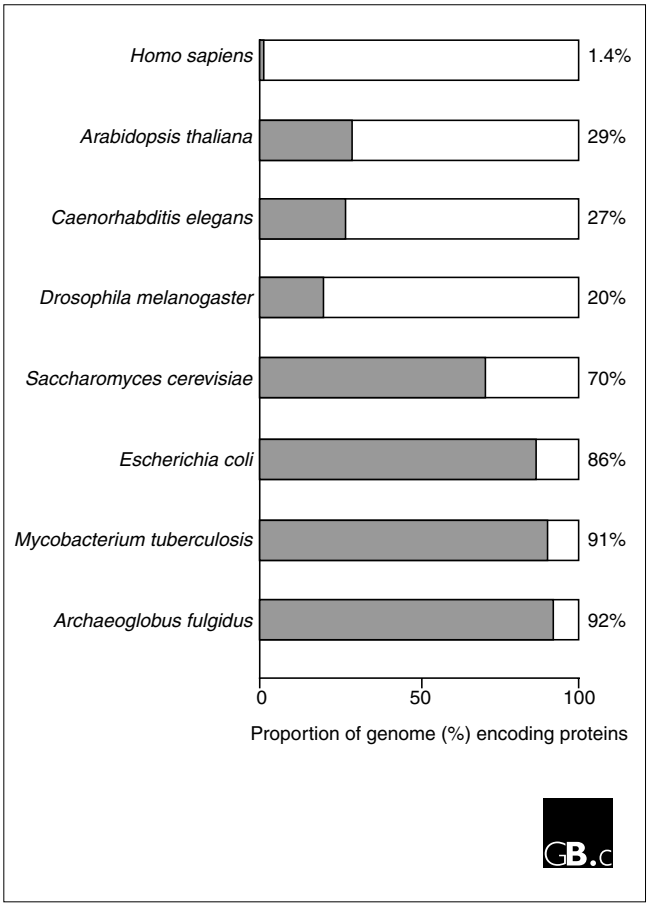

Figure 1
The percentage of protein-coding sequences (gray portions) in several eukaryotic and bacterial genomes.

tRNAs and rRNAs, many new non-protein-coding transcripts, with diverse functions, have been identified [9-11]. Non-coding RNA transcripts are heterogeneous and do not have a single specific function. Initially, the term non-coding RNA was used primarily to describe eukaryotic RNAs that are transcribed by RNA polymerase II and have a poly(A) tail at the 3′ end and a 7-methylguanosine cap structure at the 5′ end but lack a single long open reading frame (ORF). Now, this definition can be extended to cover all RNA transcripts that do not have protein-coding capacity [11], and is sometimes used to describe any piece of RNA that does not encode protein, including introns [7]. Broadly, non-protein-coding RNAs can be devided into two classes (Table 1). Housekeeping RNAs are generally constitutively expressed and required for normal function and viability of the cell; these have been the subject of many reviews [9,11,12] and are not considered further here. Regulatory non-coding RNAs, by contrast, include those that are expressed at certain stages of an organism's development or of cell differentiation, or as a response to external stimuli, and can affect the expression of other genes at the level of transcription or translation (Table 1) [13]. Here, we focus on some regulatory mechanisms in which such non-coding transcripts have been implicated.

Transcriptional regulation

The regulation of expression of particular genes usually involves specific protein transcription factors, which bind to DNA control regions (such as promoters and enhancers), thereby activating or repressing transcription of a single gene or an operon. In eukaryotes, the expression of multiple genes in specific regions of chromatin may also be regulated by alterations of chromatin structure (chromatin remodeling), facilitating or restricting access of the transcription machinery to a locus.

Dosage compensation

In most animals, males and females differ in the number of X chromosomes. The expression levels of X-chromosome genes must therefore be equalized in the two sexes, a process referred to as dosage compensation. This can be achieved either by X-chromosome inactivation in XX cells or by upregulation of the single X chromosome in XY cells. Both these mechanisms are used - by different species - and both depend upon the expression of non-coding regulatory RNAs that are key elements of the pathways leading to chromatin remodeling and hence transcriptional control.

In *Drosophila*, dosage compensation is accomplished by the two-fold enhancement of gene expression from the male X chromosome. The increased transcriptional activity of the male X chromosome depends upon specific acetylation of histone H4 on lysine 16. This modification is performed by a complex of MSL (male-specific lethal) proteins [14]. Two RNAs, *roX1* and *roX2* (RNA on X chromosome), are responsible for the assembly of the MSL protein complexes and their targeting to specific sites on the male X chromosome [15]. It has also been shown that two components of the MSL complex, the MOF (males absent on the first) protein (responsible for the acetylation of histone H4) and the MSL-3 protein, can interact directly with RNA through their chromatin-binding domains (chromodomains, found in a number of chromatin regulatory proteins). This suggests that chromodomains may be targeted to specific sites on chromosomes via interaction with non-coding RNAs [16].

In mammals, dosage compensation is achieved through transcriptional inactivation of the second X chromosome in females [17,18]. Histones in the chromatin of the inactive X chromosome are hypoacetylated and the DNA is hypermethylated, a chromatin state typical of silenced genes. Although not all details of the X-chromosome inactivation process are fully understood, it is known that it is initiated at the early stages of development. It follows the 'n-1' rule that leads to transcriptional silencing of all but one X chromosomes. It is controlled by an X-chromosome inactivation

Table I

Functional classification of non-protein-coding RNA transcripts

Housekeeping RNAs	
tRNA	Translation of genetic information
rRNA	Ribosome components; catalysis of peptide bond formation
snRNA	Pre-mRNA splicing; spliceosome components
snoRNA	RNA modification, including 2′-O-methylation and pseudouridylation
RNase P RNA	Maturation of 5′ ends of pre-tRNA
Telomerase RNA	Telomeric DNA synthesis; component of telomerase
4.5S RNA	Protein export in bacteria
7SL RNA	Protein export in eukaryotes
tmRNA	*Trans*-translation
Y RNA	Ro RNP components; function unknown
RNase MRP	Mitochondrial RNA processing
Regulatory RNAs	
Transcriptional regulators	
roX RNAs and *Xist*/*Tsix*	Chromatin remodeling associated with X-chromosome inactivation and dosage compensation in eukaryotes
H19, *IPW* and *LIT1*	Regulation of expression of imprinted genes
Post-transcriptional regulators	
DsrA, *micF*, *lin-4*, *let-7*, microRNAs, *HFE* and *LjPLP-IV*	Repression or stimulation of translation of regulated mRNAs in eukaryotic and prokaryotic cells via antisense RNA:RNA interactions
Modulators of protein function	
6S RNA, *OxyS* and *SRA*	Modulation of protein activity via RNA-protein interactions
Regulators of RNA and protein distribution	
Xlsirt and hsr-ω	Effects on localization of mRNA or pre-mRNA depending on specific subcellular location of non-coding RNA

Extensive data concerning particular classes of non-coding RNAs have been collected; details can be found in several databases [9,67-70]. Non-coding housekeeping RNAs are out of the scope of this article. Abbreviations: snRNAs, small nuclear RNAs; snoRNAs, small nucleolar RNAs; tmRNA, transfer-messenger RNA; Y RNA, Y chromosome RNA.

center (*Xic*) where the *Xist* (X inactive specific transcript) gene is located. The product of *Xist* transcription is a 17 kilobase (kb) non-coding RNA; its precise role in X-chromosome silencing is not clear. It was proposed that the process of *Xist* transcription alone might be enough to change the chromatin structure of the X chromosome in a way that would allow the binding of silencing factors. On the other hand, the accumulation of *Xist* RNA on the X chromosome that will be silenced before it is in fact inactivated suggests that its presence might be important in the deposition of silencing factors and subsequent modification of chromatin, for example by the deacetylation of histones and methylation of the promoters of X-linked genes [17-19]. It has been found that inactive X chromosomes in mouse cells have elevated levels of a specific histone isoform, macroH2A1.2, suggesting that this histone variant might be an effector of the silencing process [20]. This idea is supported by the observation that *Xist* RNA and macroH2A1.2 can form a stable ribonucleoprotein complex [21] and that the localization of macroH2A1.2 in X-chromosome chromatin depends on the expression of *Xist* [22]. The function of *Xist* RNA may be to recruit macroH2A1.2 so as to establish and maintain an inactive state [18].

Another gene associated with X-chromosome inactivation that is located within the *Xic* region is *Tsix* (antisense transcript from *Xist* locus). Expression of *Tsix* produces a 40 kb long non-coding *Tsix* RNA [23], which probably plays a role in the regulation of X-chromosome inactivation through repression of *Xist* function. It has been proposed that base-pairing between *Xist* and *Tsix* transcripts might interfere with the binding of proteins, such as macroH2A1.2, to *Xist* RNA. Another possibility is that transcription in the antisense direction inhibits synthesis of sense *Xist* transcript [18].

Genetic imprinting

Another phenomenon that somewhat resembles X-chromosome inactivation is genetic imprinting. It is a process by which modification of one of the two parental alleles of a gene results in preferential silencing of the allele from one parent. The differences in expression of paternal and maternal copies of imprinted genes are associated with differential DNA methylation or chromatin states [24]. Imprinted genes often occur in clusters and their coordinated regulation depends on the activity of an imprinting control element. In several cases, it has been demonstrated that the activity of non-coding RNA genes is essential for maintaining the imprinted status of neighboring genes. For example, the mammalian *H19* gene (encoding a non-protein-coding RNA) contains an imprinting control region that is differentially methylated and represses the paternally derived H19 allele and the maternally derived allele of the adjacent insulin-like growth factor 2 (Igf2). Similarly, the *IPW* (imprinted in Prader Willi) RNA has been suggested to function as an untranslated RNA, possibly regulating transcription in *cis* in an imprinted region associated in the Prader Willi syndrome in human and mouse. It is not known whether these transcripts themselves have any specific function: their expression might serve as an indicator of the transcriptional status of the adjacent chromosomal region. It has also been postulated that, in cases in which a non-coding RNA is transcribed from the antisense strand of the imprinted gene, the antisense RNA might participate directly in establishing or maintaining the imprinting status

(by chromatin remodeling or DNA methylation). For example, the *KvLQT1* gene (encoding a voltage-gated potassium channel) on chromosomal band 11p15 is imprinted and expressed from the maternal allele, which is disrupted in some Beckwith-Wiedemann syndrome (BWS) patients. An antisense orientation transcript within *KvLQT1*, termed *LIT1* (long QT intronic transcript 1) is expressed normally from the paternal allele, but is abnormally expressed from the maternal allele in these BWS patients and may be associated with silencing of maternally expressed genes on the same chromosome. Another possibility is that it might silence the target gene by RNA interference (RNAi), a mechanims by which the presence of double-stranded RNA induces degradation of an mRNA via a 'small-interfering' RNA intermediate produced by RNase activity [24]. Disruption of the genes encoding non-coding RNAs implicated in the regulation of imprinted genes have been found to underlie several human genetic disorders, including DiGeorge syndrome, Angelman syndrome and Prader-Willi syndrome [25,26].

The *NTT* gene (non-coding transcript in T cells), the 17 kb product of which is expressed in some activated $CD4^+$ T cells, provides another example of the possible involvement of non-coding RNAs in transcriptional regulation. Unlike the transcripts from the imprinted genes discussed above, the *NTT* gene product is produced from both parental alleles. The precise function of the *NTT* RNA is not known, but it is located close to the IFN-γR (encoding the receptor for the cytokine interferon-γ) and shows the same expression pattern as IFN-γR, suggesting that it may be involved in the regulation of IFN-γR expression [27].

Translational regulation

Many non-coding RNAs are involved in the modulation of gene expression at the post-transcriptional level. This type of regulation is widely used in prokaryotes, and recent findings suggest that it may also constitute one of the major mechanisms of gene-expression modulation in eukaryotes.

In *Escherichia coli*, non-coding RNAs have been shown to play a role in the post-transcriptional regulation of gene expression. One of the best studied and most extraordinary regulatory RNAs in *E. coli* is the *DsrA* RNA. Overexpression of this 87 nucleotide RNA reverses the transcriptional silencing that is dependent on the global repressor H-NS [28] and stimulates translation of the stress-response σ factor (RpoS) of RNA polymerase [29]. This leads to the induction of two groups of genes: those repressed by H-NS and those activated by RpoS. The levels of the H-NS and RpoS proteins are modulated at the level of translation: translational activation of RpoS depends on direct RNA:RNA interactions between the 5′ untranslated region (UTR) of the *rpoS* mRNA and the 5′ portion of *DsrA*. *DsrA* competes with a secondary structure within the *rpoS* mRNA that serves as a *cis*-acting inhibitor of translation. This model is supported by the observation that sequence complementarity between the *rpoS* mRNA and the *DsrA* RNA is essential for the stimulation of translation. RNA:RNA interactions of *DsrA* with both 5′ and 3′ portions of the ORF within the *hns* mRNA are also responsible for the repression of translation of *hns* mRNA [30,31]. It has also been noted that the *DsrA* RNA shows sequence complementary to portions of several other genes that may be post-transcriptionally regulated. The position of matching sequences in the target mRNA relative to the translation start codon might determine whether the interaction with *DsrA* has a stimulating or a repressing effect [30]. Interestingly, RpoS translation is also induced by osmotic shock, which does not result in an increase in transcription of the *DsrA* RNA. In this case, the activator function is fulfilled by *RprA* - another non-coding RNA. Although the secondary structure of *RprA* RNA is predicted to be similar to that of the *rpoS* RNA, it lacks extensive sequence complementarity to *rpoS* and the mechanism of its action is not clear [32]. Another stress-response non-coding transcript in *E. coli* is the 93 nucleotide *micF* RNA responsible for post-transcriptional control of the outer membrane porin gene *ompF*. Inhibition of *ompF* translation involves binding of the *micF* RNA to *ompF* mRNA and this interaction induces degradation of the *ompF* mRNA [33].

Studies of heterochronic mutations in *C. elegans*, which affect the timing of developmental events, identified yet another class of non-coding RNAs whose regulatory function depends on antisense interactions with target mRNAs. The products of the heterochronic *lin-4* and *let-7* genes were identified as 22 and 21 nucleotide RNAs, respectively, which are processed from 61 and 72 nucleotide precursors [34-36]. The activity of these RNAs, originally called small temporal RNAs (stRNAs), apparently depends on a sequence complementarity with the 3′ UTRs of various developmental mRNAs. Inhibition of translation is achieved after an initiation step: the targeted mRNAs are found to be associated with polyribosomes, but there is no protein product [37]. In the last year, new data have shown that *lin-4* and *let-7* RNAs are members of a new class of tiny RNAs (microRNAs) widely represented in all organisms. Computational screening of the *C. elegans* genome for non-protein-coding regions, followed by experimental verification of the transcriptional expression of these regions led to the discovery of new independent microRNA genes encoding short (approximately 65 nucleotide) precursor transcripts that can be folded into stem-loop secondary structures. These can be further processed by a specific ribonuclease to generate mature 21 to 25 nucleotide RNAs [38,39]. MicroRNAs are also expressed in *Drosophila* and mammals [40]. Some of these RNAs (for example, *mir-1* and *mir-87*) have homologs in both invertebrates and vertebrates and, in addition to controlling developmental timing, they may perform tissue-specific functions [38,40].

An antisense RNA-based mechanism has also been shown to be responsible for the regulation of the human *HFE* gene, which is implicated in iron metabolism and involved in a human inherited disorder, hereditary hemochromatosis. An antisense non-coding transcript, originating from the antisense strand of the *HFE* gene, was identified and shown to include a portion complementary to exon 1 of the sense transcript. Although there is no direct evidence for its function *in vivo*, the studies *in vitro* demonstrated that the antisense transcript represses translation of the *HFE* mRNA [41].

An antisense transcript that may function as a negative regulator of gene expression has also been identified in plants. In the legume *Lotus japonicus*, expression of the late nodulin *LjNOD16* gene is controlled by a bidirectional promoter located within an intron of the gene *LjPLP-IV* (*LjPLP-IV* encodes a phosphatidylinosiol transfer-like protein). Transcription from the opposite strand gives rise to an antisense transcript responsible for control of *LjPLP-IV* expression in root nodules, where its level is significantly lower than in flowers [42]. There are, however, no details on the mechanism by which this regulation is achieved.

Modulating protein function

Some non-coding RNAs have been shown to affect the activity of proteins directly. The association of a protein with a regulatory RNA can influence its structure as well as enzymatic and/or ligand-binding activities. One of the key regulatory RNAs working in this way in *E. coli* is 6S RNA. Because no aberrant phenotypes are associated with either null mutations or overexpression of 6S RNA, its function remained a mystery for over three decades. Recently, it has been shown that 6S RNA forms a stable complex with the σ^{70} holoenzyme of RNA polymerase [43]. This interaction modulates the activity of RNA polymerase in stationary phase (no population growth), when it may be responsible for the general reduction in transcription of σ^{70}-dependent genes or the differential use of σ^{70}-dependent promoters [43].

Like the *DsrA* RNA discussed earlier, *OxyS* RNA, which is expressed in response to oxidative stress in *E. coli*, is also a regulator of expression of the stress-response σ factor RpoS. In this case, however, translation of *rpoS* mRNA is not regulated by an antisense mechanism depending on RNA:RNA interactions but instead by a competition for the RNA-binding Hfq protein, which together with *DsrA* RNA is required for translation of *rpoS* mRNA (Figure 2) [30,44]. The *OxyS* RNA also negatively regulates translation of the *fhlA* mRNA (which encodes a transcriptional activator of genes of the formate hydrogenlyase system). In this case, *OxyS* function depends on the antisense interaction with the target mRNA that blocks the ribosome binding site [45]. In mammals, a novel non-coding RNA, the steroid receptor activator (*SRA*) RNA, has been found to function as a modulator of steroid hormone receptors. It was isolated from human and mouse cells and shown to function as a specific co-activator of several steroid receptors, including receptors for androgens, estrogens, glucocorticoids and progestins. *SRA* RNA was found to be associated in a ribonucleoprotein complex with the steroid receptor coactivator 1 (SRC-1), which is recruited by a steroid receptor. Mutations within the potential ORF of *SRA* do not affect its activity and the expression of different isoforms is cell-type-specific [46].

Regulation of RNA and protein localization

In amphibian oocytes, the correct localization of maternal mRNAs to the animal and vegetal regions determines normal embryo development. In addition to mRNAs, the vegetal cortex of *Xenopus* oocytes also contains non-coding *Xlsirt* transcripts, which contain 3 to 13 repeats of 79 to 81 nucleotide elements. *Xlsirt* RNAs are localized in the vegetal cortex at the early stages of oogenesis, and it has been proposed that they may constitute structural components of the cortex responsible for the localization of other RNAs. The importance of *Xlsirt* RNAs has been shown for the localization of the mRNA encoding Vg1, a member of the transforming growth factor β (TGF β) family of developmental signaling molecules. *Vg1* mRNA is dispersed after destruction of the *Xlsirt* RNAs with antisense oligodeoxynucleotides [47]. In *Drosophila* the nuclear transcripts of the non-coding hsr-ω were found in complexes with heterogenous nuclear RNA binding proteins (hnRNPs) in the interchromatin space. It has been suggested that the hsr-ω nuclear transcripts play a role in regulation of the trafficking and availability of hnRNPs in the nucleus [48].

Non-coding RNAs with unknown functions

The response to bone morphogenetic proteins (BMP) and osteogenic proteins (OP), members of the TGF-β superfamily that have been identified as factors responsible for the induction of bone formation *in vivo,* also seems to involve non-coding RNA transcripts. Two proteins, BMP-2 and OP-1, specifically induce transcription of the 3 kb non-coding *BORG* RNA (BMP/OP-responsive gene), which may play a key role in osteoblast differentiation although its precise function is unknown [49]. Recently it has been found that overexpression of a specific non-coding transcript from the *DD3* gene is associated with prostate cancer. Initial characterization of the gene transcripts revealed that it exists in several variants as a result of alternative splicing and alternative polyadenylation. Its expression is limited to malignant prostate cells and it does not show significant homology to any other genes. As is the case for most of non-coding RNAs, the function of this RNA is unknown and the mechanism underlying its overexpression in malignant cells has not yet been characterized [50].

Most of the data on regulatory non-coding RNAs come from the studies in animals or bacteria. Several non-coding RNA

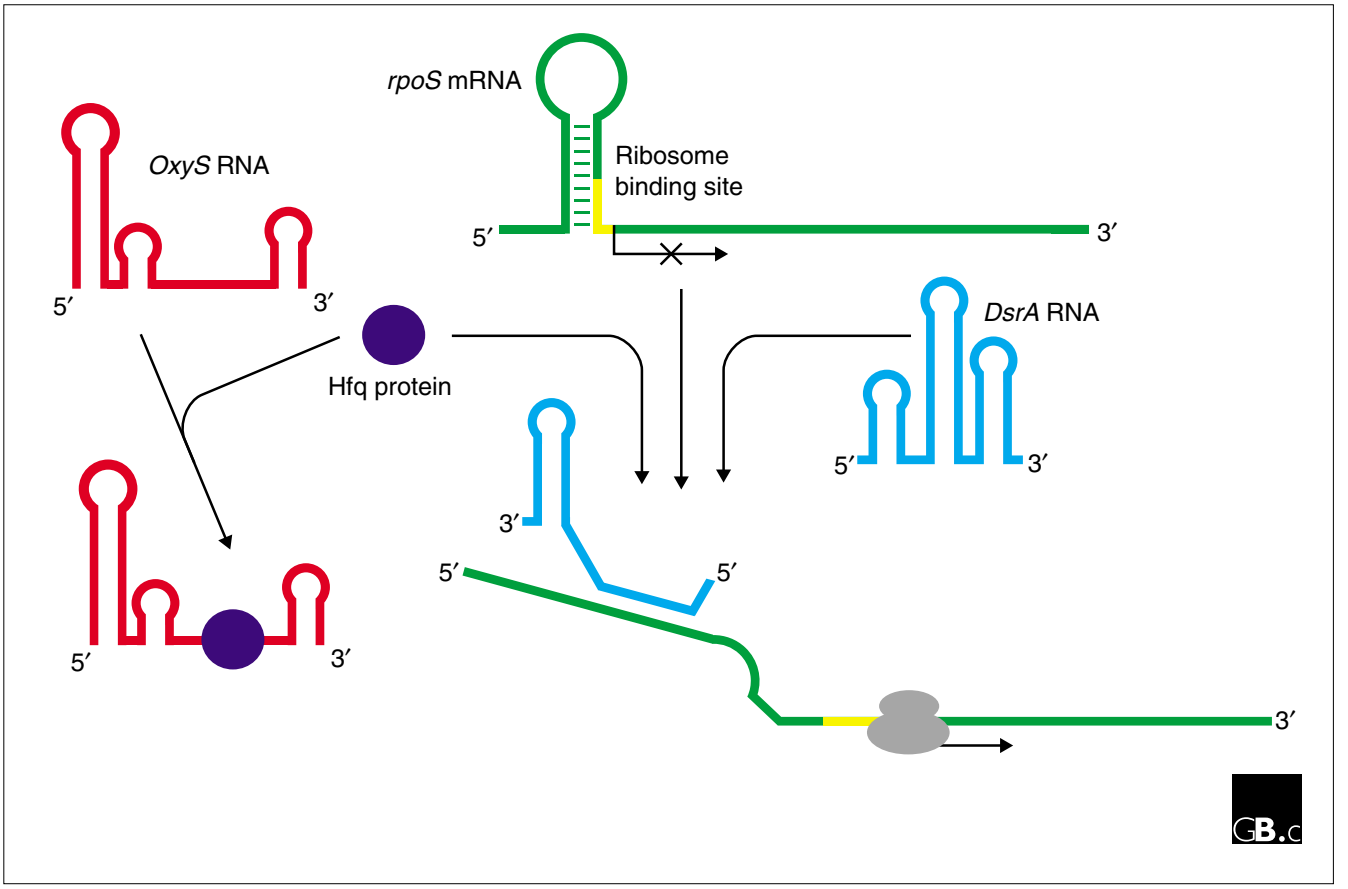

Figure 2
Translational regulation of the *E. coli* σ factor RpoS by the non-coding RNAs *DsrA* and *OxyS*. Translational activation of RpoS depends on base pairing between the *rpoS* mRNA and the *DsrA* RNA. *DsrA* competes with a secondary structure within the *rpoS* mRNA that serves as a *cis*-acting inhibitor of translation. *DsrA*-mediated translation of RpoS requires the RNA-binding protein Hfq and is negatively regulated by the *OxyS* RNA, which competes for the RNA-binding site in Hfq.

transcripts have also been isolated and partially characterized in plants, however. One of the first to be identified was the *ENOD40* RNA, which is produced in response to inoculation with the nodule-inducing bacterium *Rhizobium* or other nodulation factors. This RNA, the length of which ranges from 0.4 to 0.9 kb, has been found in several plant species [51]. The *CR20* RNA is a product of a cytokinin-responsive gene that is repressed in response to cytokinins (plant hormones) or stress conditions and was first isolated from cucumber and later reported in several other plant species [52]. Another hormonally regulated transcript is *GUT15* (gene with unstable transcript 15) from tobacco [53]. The function of the *CR20* and *GUT15* transcripts is unknown, but their hormonal regulation and low stability suggest that they may play regulatory roles. *Medicago truncatula Mt4* RNA and tomato *TPSI1* represent another family of plant non-coding transcripts upregulated by phosphate starvation [54,55]. Members of this family show a very high degree of nucleotide sequence conservation, but there is no evidence that they are translated into protein products.

Searching genomic sequences for non-coding RNAs

Most gene-finding algorithms are designed to look for protein-coding sequences, which can be more readily identified than non-coding RNAs by virtue of their ORFs, polyadenylation signals, conserved promoter regions or splice-site signals. Because it was assumed that non-coding RNAs of interest would have stable secondary structures, early ideas about how to identify RNA-coding genes concentrated on secondary structure prediction by energy minimization [56]. A modified approach, using stochastic context-free 'grammar' for RNA structure prediction, has been used to screen several genomic sequences [57]. The results of these studies [56,57] led to the conclusion that the secondary structures of genuine non-protein-coding RNAs cannot be distinguished from the structures predicted for random RNA sequences, so these methods are unusable for predicting non-protein-coding genes.

Currently, the identification of RNA-encoding genes in genomic sequences is based on structural or sequence

homologies. There are efficient programs that search for tRNA or small nucleolar RNA (snoRNA) genes by using conserved structural elements or sequences inferred from the analysis of known RNAs, for example [58,59]. For a global search approach that would work for all functional RNAs, one would have to assume that there are significant signals in all protein- and RNA-coding sequences that can be used to distinguish them from regions of the genome that are transcriptionally inactive. Methods using computational neural networks and support vector machines have been used to extract common sequence features and structural elements from known RNAs, for example; these parameters were then used to screen eubacterial and archaeal genomes [60]. The results showed that RNA-coding sequences do, in fact, contain information that can be used for accurate gene finding.

The wealth of genomic sequences now available from a variety of organisms allows comparative sequence analysis, which can potentially help to identify important sequences that cannot be detected by analysis of individual genomes. Such comparisons should distinguish structural RNAs from other conserved sequences, assuming that structural RNAs show compensatory mutations consistent with their secondary structure [61]. Comparison of the intergenic regions of the *E. coli* genome with the genomes of five other enterobacteria and analysis of the resulting pairwise BLASTN sequence alignments using the QRNA program, which searches for conserved RNA structures, identified 275 potential non-coding RNA sequences [61]. Subsequent experiments confirmed that some of these sequences are in fact functional non-coding RNA genes [62,63].

Another approach used to find novel non-coding RNA genes is a combination of computational and experimental methods. A search in yeast for RNA polymerase III promoters, typically found in small RNA genes (such as tRNA genes), and analysis of the expression from the sequence 'gaps' between the predicted ORFs, led to the identification of novel non-coding RNA transcripts as well as of RNAs containing small ORFs [64]. The identification of several non-coding RNA genes in plants, the expression of which is modulated by biotic and abiotic stress conditions [13], as well as the observation that RNA can be transported over long distances by the phloem sieve tubes [65], suggests that RNA may be widely employed as a signaling molecule in plants. Because all of the plant non-coding RNAs described so far have mRNA characteristics, such as poly(A) tails and caps, expressed sequence tag (EST) sequences from *Arabidopsis thaliana* were systematically screened, and 19 clones that probably function as non-protein-coding RNAs were identified. These clones are apparently plant-specific transcripts with no homologs outside the plant kingdom [66].

In conclusion, the discovery of non-coding regulatory RNAs and the variety of molecular phenomena in which such RNA molecules have been implicated suggest that non-coding RNAs may play key roles in the overall molecular organization of organisms. From the point of view of cell economy, RNA is well suited to be a signaling molecule: RNA can be synthesized in response to a particular stimulus and can then be rapidly destroyed without the necessity of costly protein synthesis. It has also been postulated that regulatory RNA molecules could originate from the introns of protein-coding genes as functional by-products [6,7]. The growing number of new, functional, non-coding RNAs shows that to fully understand the molecular mechanisms in a cell we have to go beyond the predicted proteome when analyzing genomic sequences.

References

1. Hastings ML, Krainer AR: **Pre-mRNA splicing in the new millennium.** *Curr Opin Cell Biol* 2001, **13:**302-309.
2. Graveley BR: **Alternative splicing: increasing diversity in the proteomic world.** *Trends Genet* 2001, **17:**100-107.
3. Levy M, Ellington AD: **RNA world: catalysis abets binding, but not *vice versa*.** *Curr Biol* 2001, **11:**R665-R667.
4. Ban N, Nissen P, Hansen J, Moore PB, Steitz TA: **The complete atomic structure of the large ribosomal subunit at 24 Å resolution.** *Science* 2000, **289:**905-920.
5. Caprara MG, Nilsen TW: **RNA: versatility in form and function.** *Nat Struct Biol* 2000, **7:**831-833.
6. Mattick JS: **Non-coding RNAs: the architects of eukaryotic complexity.** *EMBO Rep* 2001, **2:**986-991.
7. Mattick JS, Gagen MJ: **The evolution of controlled multitasked gene networks: the role of introns and other noncoding RNAs in the development of complex organisms.** *Mol Biol Evol* 2001, **18:**1611-1630.
8. Shabalina SA, Ogurtsov AY, Kondrashov VA, Kondrashov AS: **Selective constraint in intergenic regions of human and mouse genomes.** *Trends Genet* 2001, **17:**373-376.
9. Erdmann VA, Barciszewska MZ, Szymanski M, Hochberg A, de Groot N, Barciszewski J: **The non-coding RNAs as riboregulators.** *Nucleic Acids Res* 2001, **29:**189-193.
10. Eddy SR: **Noncoding RNA genes.** *Curr Opin Genet Dev* 1999, **9:**695-699.
11. Eddy SR: **Non-coding RNA genes and the modern RNA world.** *Nat Rev Genet* 2001, **2:**919-929.
12. Wassarman KM, Zhang A, Storz G: **Small RNAs in *Escherichia coli*.** *Trends Microbiol* 1999, **7:**37-45.
13. Erdmann VA, Barciszewska MZ, Hochberg A, de Groot N, Barciszewski J: **Regulatory RNAs.** *Cell Mol Life Sci* 2001, **58:**960-977.
14. Smith ER, Pannuti A, Gu W, Steurnagel A, Cook RG, Allis CD, Lucchesi JC: **The *Drosophila* MSL complex acetylates histone H4 at lysine 16, a chromatin modification linked to dosage compensation.** *Mol Cell Biol* 2000, **20:**312-318.
15. Meller VH, Gordadze PR, Park Y, Chu X, Stuckenholz C, Kelley RL, Kuroda MI: **Ordered assembly of *roX* RNAs into MSL complexes on the dosage-compensated X chromosome in *Drosophila*.** *Curr Biol* 2000, **10:**136-143.
16. Akhtar A, Zink D, Becker PB: **Chromodomains are protein-RNA interaction modules.** *Nature* 2000, **407:**405-409.
17. Avner P, Heard E: **X-chromosome inactivation: counting, choice and initiation.** *Nat Rev Genet* 2001, **2:**59-67.
18. Maxfield Boumil R, Lee JT: **Forty years of decoding the silence in X-chromosome inactivation.** *Hum Mol Genet* 2001, **10:**2225-2232.
19. Clemson CM, McNeil JA, Willard H, Lawrence JB: ***XIST* RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure.** *J Cell Biol* 1996, **132:**259-275.
20. Constanzi C, Pehrson JR: **MacroH2A1 is concentrated in the inactive X chromosome of female mammals.** *Nature* 1998, **393:**599-601.
21. Gilbert SL, Pehrson JR, Sharp PA: ***XIST* RNA associates with specific regions of the inactive X chromatin.** *J Biol Chem* 2000, **275:**36491-36494.

22. Csankovszki G, Panning B, Bates B, Pehrson JR, Jaenisch R: **Conditional deletion of *Xist* disrupts histone microH2A localization but not maintenance of X-chromosome inactivation.** *Nat Genet* 1999, **22:**323-324.
23. Lee JT, Davidow LS, Warshawsky D: ***Tsix*, a gene antisense to *Xist* at the X-inactivation center.** *Nat Genet* 1999, **21:**400-404.
24. Reik W, Walther J: **Genomic imprinting: parental influence on the genome.** *Nat Rev Genet* 2001, **2:**21-32.
25. Sutherland H, Wadey R, McKie JM, Taylor C, Atif U, Johnstone KA, Halford S, Kim UJ, Goodship J, Baldini A, Scambler PJ: **Identification of a novel transcript disrupted by a balanced translocation associated with DiGeorge syndrome.** *Am J Hum Genet* 1996, **59:**23-31.
26. Meguro M, Mitsuya K, Nomura N, Kohda M, Kashiwagi A, Nishigaki R, Yoshioka H, Nakao M, Oishi M, Oshimura M: **Large-scale evaluation of imprinting status in the Prader-Willi syndrome region: an imprinted direct repeat cluster resembling small nucleolar RNA genes.** *Hum Mol Genet* 2001, **10:**383-394.
27. Liu AY, Torchia BS, Migeon BR, Siliciano RF: **The human *NTT* gene: identification of a novel 17-kb noncoding nuclear RNA expressed in activated CD4$^+$ T cells.** *Genomics* 1997, **39:**171-184.
28. Sledjeski D, Gottesman S: **A small RNA acts as an antisilencer of the H-NS-silenced *rcsA* gene of *Escherichia coli*.** *Proc Natl Acad Sci USA* 1995, **92:**2003-2007.
29. Sledjeski DD, Gupta A, Gottesman S: **The small RNA, *DsrA*, is essential for the low temperature expression of RpoS during exponential growth in *Escherichia coli*.** *EMBO J* 1996, **15:**3993-4000.
30. Lease RA, Cusick ME, Belfort M: **Riboregulation in *Escherichia coli:* DsrA acts by RNA:RNA interactions at multiple loci.** *Proc Natl Acad Sci USA* 1998, **95:**12456-12461.
31. Lease RA, Belfort M: **Riboregulation by *DsrA* RNA: transactions for global economy.** *Mol Microbiol* 2000, **38:**667-672.
32. Majdalani N, Chen S, Murrow J, St John K, Gottesman S: **Regulation of RpoS by a novel small RNA: the characterization of RprA.** *Mol Microbiol* 2001, **39:**1382-1394.
33. Delihas N, Forst S: ***MicF*: an antisense RNA gene involved in response of *Escherichia coli* to global stress factors.** *J Mol Biol* 2001, **313:**1-12.
34. Lee RC, Feinbaum RL, Ambros V: **The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*.** *Cell* 1993, **75:**843-854.
35. Reinhart BJ, Slack FJ, Basson M, Pasquinelli AE, Bettinger JC, Rougvie AE, Horvitz HR, Ruvkun G: **The 21-nucleotide *let-7* RNA regulates developmental timing in *Caenorhabditis elegans*.** *Nature* 2000, **403:**901-906.
36. Moss EG: **Non-coding RNAs: lightning strikes twice.** *Curr Biol* 2000, **10:**R436-R439.
37. Olsen PH, Ambros V: **The *lin-4* regulatory RNA controls developmental timing in *Caenorhabditis elegans* by blocking LIN-14 protein synthesis after the initiation of translation.** *Dev Biol* 1999, **216:**671-680.
38. Lau NC, Lim LP, Weinstein EG, Bartel DP: **An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*.** *Science* 2001, **294:**858-862.
39. Lee LC, Ambros V: **An extensive class of small RNAs in *Caenorhabditis elegans*.** *Science* 2001, **294:**862-864.
40. Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T: **Identification of novel genes coding for small expressed RNAs.** *Science* 2001, **294:**853-858.
41. Thenie AC, Gicquel IM, Hardy S, Ferran H, Fergelot P, Le Gall J-Y, Mosser J: **Identification of an endogenous RNA transcribed from the antisense strand of the *HFE* gene.** *Hum Mol Genet* 2001, **10:**1859-1866.
42. Kapranov P, Routt SM, Bankaitis VA de Bruijn FJ, Szczyglowski K: **Nodule-specific regulation of phosphatidylinositol transfer protein expression in *Lotus japonicus*.** *Plant Cell* 2001, **13:**1369-1382.
43. Wassarman KM, Storz G: **6S RNA regulates *E. coli* RNA polymerase activity.** *Cell* 2000, **101:**613-623.
44. Zhang A, Altuvia S, Tiwari A, Argaman L, Hengge-Aronis R, Storz G: **The OxyS regulatory RNA represses rpoS translation and binds the Hfq HF-I protein.** *EMBO J* 1998, **17:**6061-6068.
45. Altuvia S, Zhang A, Argaman L, Tiwari A, Storz G: **The *Escherichia coli OxyS* regulatory RNA represses fhlA translation by blocking ribosome binding.** *EMBO J* 1998, **17:**6069-6075.
46. Lanz RB, McKenna NJ, Onate SA, Albrecht U, Wong J, Tsai SY, Tsai MJ, O'Malley BW: **A steroid receptor coactivator, SRA, functions as an RNA and is present in an SRC-1 complex.** *Cell* 1999, **97:**17-27.
47. Kloc M, Etkin LD: **Delocalization of *Vg1* mRNA from the vegetal cortex in *Xenopus* ooctes after destruction of *Xlsirt* RNA.** *Science* 1994, **265:**1101-1103.
48. Prasanth KV, Rajendra TK, Lal AK, Lakhotia SC: **Omega speckles - a novel class of nuclear speckles containing hnRNPs associated with noncoding hsr-omega RNA in *Drosophila*.** *J Cell Sci* 2000, **113:**3485-3497.
49. Takeda K, Ichijo H, Fujii M, Mochida Y, Saitoh M, Nishitoh H, Sampath TK, Miyazono K: **Identification of a novel bone morphogenetic protein-responsive gene that may function as a noncoding RNA.** *J Biol Chem* 1998, **273:**17079-17085.
50. Bussemakers MJ, van Bokhoven A, Verhaegh GW, Smit FP, Karthaus HF, Schalken JA, Debruyne FM, Ru N, Isaacs WB: ***DD3*: a new prostate-specific gene, highly overexpressed in prostate cancer.** *Cancer Res* 1999, **59:**5975-5979.
51. Staehelin C, Charon C, Boller T, Crespi M, Kondorosi A: ***Medicago truncatula* plants overexpressing the early nodulin gene *enod40* exhibit accelerated mycorrhizal colonization and enhanced formation of arbuscules.** *Proc Natl Acad Sci USA* 2001, **98:**15366-15371.
52. Teramoto H, Toyama T, Takeba G, Tsuji H: **Noncoding RNA for *CR20*, a cytokinin-repressed gene of cucumber.** *Plant Mol Biol* 1996, **32:**797-808.
53. van Hoof A, Kastenmayer JP, Taylor CB, Green PJ: **GUT15 cDNAs from tobacco (AC No. U84972) and *Arabidopsis* (AC No. U84973) correspond to transcripts with unusual metabolism and a short conserved open reading frame.** *Plant Physiol* 1997, **113:**1004.
54. Burleigh SM, Harrison MJ: **Characterization of the *Mt4* gene from *Medicago truncatula*.** *Gene* 1998, **216:**47-53.
55. Mukatira UT, Liu C, Varadarajan DK, Raghothama KG: **Negative regulation of phosphate starvation-induced genes.** *Plant Physiol* 2001, **127:**1854-1862.
56. Le S-J, Chen J-H, Maizel J: **A program for predicting significant RNA secondary structures.** *Comput Appl Biosci* 1988, **4:**153-159.
57. Rivas E, Eddy SR: **Secondary structure alone is generally not statistically significant for the detection of noncoding RNAs.** *Bioinformatics* 2000, **16:**583-605.
58. Lowe T, Eddy SR: **tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequences.** *Nucleic Acids Res* 1997, **25:**955-964.
59. Lowe T, Eddy SR: **A computational screen for methylation guide snoRNAs in yeast.** *Science* 1999, **283:**1168-1171.
60. Carter RJ, Dubchak I, Holbrook SR: **A computational approach to identify genes for functional RNAs in genomic sequences.** *Nucleic Acids Res* 2001, **29:**3928-3938.
61. Rivas E, Klein RJ, Jones TA, Eddy SR: **Computational identification of noncoding RNAs in *E. coli* by comparative genomics.** *Curr Biol* 2001, **11:**1369-1373.
62. Wassarman KM, Repoila F, Rosenow C, Storz G, Gottesman S: **Identification of novel small RNAs using comparative genomics and microarrays.** *Genes Dev* 2001, **15:**1637-1651.
63. Argaman L, Hershberg R, Vogel J, Bejerano G, Wagner GEH, Margalit H, Altuvia S: **Novel small RNA-encoding genes in the intergenic regions of *Escherichia coli*.** *Curr Biol* 2001, **11:**941-950.
64. Olivas WM, Muhlard D, Parker R: **Analysis of the yeast genome: identification of new non-coding and small ORF-containing RNAs.** *Nucleic Acids Res* 1997, **25:**4619-4625.
65. Lucas WJ, Yoo B-C, Kragler F: **RNA as a long-distance information macromolecule in plants.** *Nat Rev Mol Cell Biol* 2001, **2:**849-857.
66. MacIntosh GC, Wilkerson C, Green PJ: **Identification and analysis of *Arabidopsis* expressed sequence tags characteristic of non-coding RNAs.** *Plant Physiol* 2001, **127:**765-776.
67. ***Nucleic Acids Research* database categories list**
[http://www3.oup.co.uk/nar/database/c]
68. **Database of non-coding RNAs**
[http://biobases.ibch.poznan.pl/ncRNA/]
69. **Small RNA database**
[http://mbcr.bcm.tmc.edu/smallRNA/smallrna.html]
70. **Database of plant ncRNAs**
[http://www.prl.msu.edu/PLANTncRNAs/database.html]