RNA polymerase II dependent genes that do not code for protein

S.C. Lathia
Genetics Laboratory, Department of Zoology, Banaras Hindu University, Varanasi 221 005

Received 14 August 1994; accepted 21 December 1995

In recent years more and more examples of RNA polymerase II dependent non-coding trans-
scripts have been described. Although these have frequently been ignored as "selfish DNA ele-
ments", it is becoming increasingly clear that many, if not all, of them have very important biological
roles. Examples of such "genes" from Drosophila, mammals, other vertebrates, yeast etc. are consid-
ered. Although the specific mechanisms through which these non-coding transcripts function in the
cell are not clear, comparisons reveal certain common themes, particularly the importance of secon-
dary structures, rather than the primary base sequence of these transcripts. While some of these
transcripts may function as ribozymes or as antisense regulators, most others may function more di-
rectly through their specific protein-binding properties. Since RNA is believed to be the first "living"
macromolecule, it is very likely that some genes even today function only through this class of molecules. It
is expected that instead of being ignored as examples of "selfish DNA", a more positive search for
their functions will help unravel the significance of this novel class of genes.

Introduction

Eukaryotes have three principal RNA polymer-
eses, each earmarked for transcribing specific class of genes. The RNA polymerase I is responsible
for transcribing DNA sequences coding for the major ribosomal RNAs while the RNA polymerase
III is used to transcribe the SS ribosomal RNA, the various tRNAs and the small nuclear RNAs (snRNA). The transcripts produced by RNA polymerases I and III are not translated but are involved in processing and transcribing the transcripts made by the RNA polymerase II which transcribes all the protein-coding genes dis-
persed throughout chromosomal DNA. The RNA polymerase II transcribed genes share many fea-
tures in common with regard to their transcriptional regulation, post-transcriptional processing of the primary transcripts (intragenic nuclear RNA or hnRNA) and finally the transport of the processed messenger RNA (mRNA) to the cyto-
plasm for translation into their specific polypept-
ide product. The central dogma of molecular bi-
ology has altered such an important rule to the transla-
tional activities of mRNA that any RNA polymerase II transcribed gene is expected to have a protein product. Those that seemed to not code for a protein have often been put aside as "selfish DNA". However, over the years more and more instances of genes that are transcribed by RNA polymerase II, whose products may display typical post-transcriptional processing events and are yet not coding for any protein, have been discovered in diverse organisms. This review will consider some of the known examples of "non-
coding" genes.

Y-Chromosome loops or the fertility genes in Drosophila males

Ever since the classical studies of Bridges on sex determination in Drosophila, it has been known
that the Y-chromosome is not essential for sex determination but is essential for fertility of male flies. However, whole of the Y-chromosome is condensed, heterochromatic and genetically "in-
ert" and, therefore, its essential role in male ferti-
ility remained enigmatic although conventional genet-
ic studies did identify a number of Y-linked "fertility genes" having specific effects on sper-
matogenesis. Meyer and his group12 showed that
during the primary spermatocyte stage in D. hy-
brid, the Y-chromosome opents up into very large microscopically visible and transcriptionally very active "lampbrush" loops which are essential for normal differentiation of spermatids into mature spermatozoa. In D. hydei, six distinctive loops with characteristic morphology and transcrip-
tion patterns are present while in D. melanogaster, the Y-chromosome loops in primary spermatoc-
ocytes are not so distinct but are comparable in their general organization to that in D. hydei.
Y-Linked mutations that affect fertility have a complex relationship with the Y-chromosome-associated loops. The most remarkable feature of the Y-chromosomal linked fertility genes and the lampbrush chromosomes is the enormous size of the transcription units and the nature of DNA associated with them: the Y-chromosomal transcription units in primary spermatocytes are as large as 40 000 kb in D. melanogaster and between 260 to 1500 kb in D. hydei. Bulk of these huge transcription units is comprised of simple, satellite and of functional or degenrating transposable elements. None of these have any substantial open reading frames. The only conventional protein coding gene so far known to be present on the D. melanogaster Y-chromosome is the one that produces a male-specific beta-heavy chain of dynin interlobular minor. The kl-5 complementation unit of the Y-chromosomes, at which the dynin gene maps, is estimated to be about 1200 kb long while the dynin transcript is estimated to be only about 34 kb. The physical arrangement of the dynin mRNA coding region with rest of the transcription unit is not well understood; it has been speculated that the dynin-coding region may actually be outside of the kl-5 transcription unit or may be scrambled in its 1500 kb region.

All the known Y-chromosomal transcription units, active only during the primary spermatocyte stage, share a few common genetic features. These include (a) productions of very large sized transcripts that remain restricted to the nucleus, (b) the transcription units comprise essentially of simple repetitive and retrotransposon elements which show remarkable sequence diversity between related species, (c) the base sequences of these transcription units are full of stop codons in all possible reading frames and thus do not seem to code for any protein (except for the above noted dynin gene) and (d) the various transcripts bound to specific proteins and are responsible for the characteristic shape and size of the different Y-chromosomal lampbrush loops. It appears that during the evolution of these male fertility genes, many retrotransposons got incorporated within the lampbrush transcription units. Interestingly such insertions do not appear to have affected the functioning of these genes. Thus the male fertility gene Q in the lampbrush loops called "Qnoises" has a large number of gypsy retroposons. Sequence analysis of these gypsy elements reveals that all of them are transcribed and those lost those sequences that may interfere with transcriptional continuity along the loop. Transcription of these retroposons is required but not sufficient for the function of fertility genes like the Q gene (Qnoises) of D. hydei. It is also interesting to note that the many tandem repeats of retroposons that are present in the loop transcription units show identical sequence orientation although outside the loop boundary the same retroposons show random arrangement. What is the function of these unusually large and apparently untranslatable RNA's? The Y-linked fertility genes and the lampbrush loops are essential for spermiogenesis to proceed normally. The discovery of at least one protein-coding genes (Dynin) on the Y-chromosome of D. melanogaster has encouraged hopes of finding more such protein-coding genes associated with the other fertility genes such hopes have further encouraged suggestions that the huge transcription units comprising of satellite sequences and transposons, have evolved because the fertility genes are "the ultimate heavens for selfish gene elements". However, an alternative and a challenging hypothesis suggests these transcripts to have more of structural and/or regulatory role by providing, substrates to which other macromolecules may bind.

Other non-coding genes expressed in germ line cells of male Drosophila.

Besides the above considered Y-chromosomal loops, a few other non-coding "genes" active during spermatogenesis in Drosophila are also worthy. One of the them is the enigmatic "crystal" (or Suppressor of Sclerotome-Scidue) set of genes D. melanogaster males without a Y-chromosome (XO males) show presence of needle-like or star-like inclusions in their primary spermatocytes, a high level of non-disjunction of mitotic chromosomes, abnormal distribution of organelles in meiotic and events of meiotic drive. A Y-chromosomal linked locus, the crystal or cry2 (also named Suppressor of Sclerotia, Sli[Sc]) and an X-linked locus, the Staude I or Ste2, are involved in this phenotype of XO male D. melanogaster. These two X- and Y-linked loci contain arrays of partially homologous, randomly repeated sequences. The normal functioning of cry or Sli[Sc] locus requires a critical number of subunits rather than physical integrity of the whole array of tandem repeats. The Ste allele of Sli[Sc] locus causes production of needle-like inclusions in the whole population of male Drosophila.
xpermatogenesis while the Ste allele causes star-like inclusions13. The Steu locus codes for a protein which shares homology with the beta subunit of casein kinase II12. The Steu loci on the Y-chromosome suppresses activity of the Steu locus on the X-chromosome so that in XY individuals, only a few stellite transcripts are made and even these are not properly spliced and processed but in the absence of Y-chromosome or in the presence of Y-chromosome deficient for the Steu loci, the level of stellite transcripts in spermatogenesis is significantly elevated resulting in the formation of needle-like star-like inclusions and to consequently affect disjunction and segregation of homologous chromosomes and cytoplasmic organelles like mitochondria7,8,9. The Steu loci besides affecting homology with the Ste locus, also share sequence homology with the Ste-U family of transposable4. The 2800 bp repeat unit of the SteU loci has been shown to consist of a region of homology to the Ste locus, a Y-specific Ad-rich segment and a mobile element 1360 inserted in the Ste sequence11. It is not known if the SteU transcripts code for a protein but this appears unlikely in view of its structure. The suppression of Ste activity in the absence of Ste in the sequence of SteU has been suggested as well as to be due to a mutual competition for a limiting set of transcription and splicing factors so that in the absence of SteU, these will be available to Ste for an abundant production of the stellite protein. This mutual inactivation of these two genes has also been considered an intriguing type of selfish genetic systems since related species of Droso-
phila do not have Ste or SteE loci. However, the important point to note is that in this case also, the SteU gene has very vital effect on the organism without possibly having a typical protein product.

Another gene active in the male germ line of D. melanogaster, but apparently not having a protein product, is the recently identifiedMel-40 sequence (Male specific transcript 40) located on the section 40 of the right arm of chromosome 2: the Mel-40 sequence is organized as tandemly arrayed 1.4 kb repeat unit with transcripts limited to male in male germ line, the longest possible open reading frame is 43 amino acid long without homology to any known polypeptides; this gene was derived in all strains of D. melanogaster tested although other species of Drosophila did not show its presence18. A function for this gene is yet to be found.

Among the diverse variety of transposons that make up the bulk of Y-chromosome in Drosophi-
la is the microRNA family. These retrotransposons are present both on the autosomes, X-chromosome and the Y-chromosome of D. hydei as well as D. melanogaster. Lamoureux et al. have reported that in maleic cells of male D. hydei, in addition to full length transcripts, microRNA also encodes antisense transcripts complementary to the reverse transcriptase and RNase H coding region; these antisense transcripts are present only in male germ line since they are produced from a testis specific promoter. Furthermore, while most of these transcripts are present as part of full RNA molecules because of their location on the long-patch loop forming sites, the testis and microRNA transcripts are 1.0 and 1.6 kb long. It appears that these full length antisense transcripts have an important role in the control of microRNA encoded reverse transcriptase protein in male germ line8,9.

The 3SD or the HSRW gene of Drosophila

One of the first non-protein coding genes to be characterized in some detail is the 3SD or the hsr3 gene of D. reaumuri and its homologues in other species of Drosophila. Transcriptionally, this is one of the most active genes following heat shock.12,13. That this gene was different from other heat shock genes was revealed by its singular in- ducibility in polytene cells treated with a variety of agents like benzamide, colchicine etc. (see reviews in ref. 32-34). Since no new protein were induced when this locus was selectively activated by benzamide, Lakhota and Makker suggest that this gene does not code for any protein; this was confirmed when this gene and its homo- logues in other species were cloned and se- quenced18-20. Sequence analysis revealed a remarkably conserved overall organisation of this gene but equally remarkably divergence in the sequence in different Drosophila species. In all species, the locus spans more than 10 kb and includes two exons and an intron in the proximal 1.9 to 2.0 kb followed (on the 3' end) by a long stretch of tandem arrays of repeat units unique to this locus. The base sequence of the unique as well as of the repeat units is not strongly con- served between species except for certain small regions at the exon-intron junctions and for a 9 bp motif in the repeat units20,21. Although the repeat units do not share homology between different species, all the repeats in a tandem array in a species are highly homogenous and maintain a certain minimum and maximum length21. The hsr3 locus produces two primary nuclear trans- scripts of >10 kb and ~1.9 kb size, respectively,
the 1.9 kb transcript spanning the two exons and one intron is spliced to produce the 1.2 kb poly A containing transcript. The 1.2 kb transcript has only one very short ORF but the amino acid sequence coded by this ORF is not conserved in different species19,20,24,25. The relative abundance of the three transcripts produced by this locus depends upon the nature of inducing44,44. Unlike most genes, the spliced out intron in this case is highly stable44,44. Besides being induced by the different inducers, this locus is also developmentally active in most tissue types of embryo, larva, prepup and adult44 and although without a protein coding function, it is essential for survival of flies44 and also for development of normal thorax-dermice44. Hsp 63 is known to bind to the heat shock induced 65D locus19 and recent observations in our laboratory suggest that the lethality due to deficiency of this locus is enhanced in S18S5 mutant heterozygotes44.

A series of studies in our laboratory [reviewed in refs 32-34] showed that the 93D locus affected synthesis and/or turnover of the hsp 70 and αf repeat heat induced but non-translatable, see below - transcripts. The rates of synthesis and/or turnover from the site of synthesis of the 87A-type and 87C-type hsp70 and the αf transcripts varied in relation to the specific profile of the 93D transcripts present or synthesized in response to a given condition of heat shock24. In this context, it is interesting to note that the five copies of hsp 70 genes in D. melanogaster, present at the 87A (2 copies) and 87C (3 copies) sites, share nearly identical coding and the 5' upstream regulatory regions but their 3' untranslated regions (5' UTR) show considerable divergence. Shrama and Lakhin17 suggested that the differing 5' UTRs may target the hsp70 mRNA to different cellular compartments and that the 93D transcripts have a role in this process.

It is obvious that the 93D locus in D. melanogaster and its homologue in other species have important functions to perform in almost all tissue types during normal development as well as under various conditions of cellular stress44,20,24. It has been suggested that one of the functions of the cytoplasmic 1.2 kb transcript is to “monitor the health” of translational machinery while the nuclear >10 kb transcript may be involved in synthesis and turnover/transport of other transcripts like the hsp7024,24.

The αf repeats of D. melanogaster

This is an interesting family of repetitive sequences that is present at several locations in the genome of D. melanogaster with some of them being heat shock inducible. The 87C site, locus for 3 copies of 129P70 genes, also harbours about 10-14 copies of αf repeats44. In addition, the αf repeats are also present at the heterochromatinic chromocentre but these are not heat inducible. Those at the 87C site are heat inducible due to their being associated with sequence elements (the α elements) that are identical to the hsp 70 promoter region47,44. The αf units that are immediately downstream of the α elements at the 87C site are induced by heat shock to produce multiple poly-A transcripts of 2.5, 1.8, 1.4 and 1.1 kb sizes44. None of these appear to code for any protein47,44. The suggestion that these sequence are one more example of “selfish DNA” has gained support from observations that deletion of these sequences from the 87C site has no deleterious effect and that a sibling species, D. simulans, does not carry any heat inducible αf repeats at the 87C or at any other site44,44. Nevertheless as noted above, a series of studies in our laboratory has shown that the non-protein coding heat shock loci at 93D has specific effect on transcription of the αf sequences during heat shock24,32,34,44,47. Collectively treatment which reduces transcription at the 93D locus (see above) also leads to an increase in the level of αf transcripts at the 87C site44. Significance of these interactions remains unknown.

Non-protein-coding genes in mammals

In recent years, a number of genes associated with specific loci but apparently not coding for any protein have been identified in different mammalian genomes. These are briefly considered below.

The Xist gene

All mammals show inactivation of one of the two X-chromosomes in somatic cells of females to achieve dosage compensation of X-chromosome linked genes in males and females22. While the paternal X-chromosome is preferentially inactivated in Mus musculus, the inactivation of one of the two X-chromosomes in catherina is generally random in different somatic cells but once inactivated, the same X-chromosome continues to remain inactive in all cell generations. This inactivation affects condensation, transcription and replication of the entire chromosome and is apparently regulated by a single cis-acting centre, the X-inactivation centre [termed Xic in human and Xic in mouse], which is not only responsible for the induction of inactivation but also for its spread.
the entire chromosomal length (reviewed in refs 58-60). A broad breakthrough in understanding of this whole chromosome inactivation process was the cloning and characterization of a human as well as mouse gene that appeared to correspond to the Xc or XxchA11,12. This gene, termed XIST (human) or Xist (mouse) is an inactive specific transcript, has attracted considerable attention not only for its remarkable role in inactivation of a whole chromosome but also for the way it achieves this role. The human Xist mRNA is 17kb long while the mouse Xist mRNA is 15 kb long but none of these appears to code for any protein24,13 and in both cases, the transcripts are made only by the inactive X chromosome, the allele on the active X chromosome remaining completely silent. Xist transcripts are exclusively nuclear and appear to remain associated with the Barr body, which represents the inactive X-chromosome in inactivated nuclei.6 Evidence for involvement of Xist transcripts in inactivation of the X-chromosome appears complete since the appearance of these transcripts shows an absolute parallel with the pattern of X-inactivation. In mouse, humans and other eutherians, the first sign of X-inactivation is seen in extra-embryonic trophoblast cells and primitive endoderm lines with exclusive inactivation of the paternal derived X-chromosome in all cells.14 X-inactivation in embryonic lines occurs later and this is probably with respect to the parental origin of the X-chromosome. Paternal X-chromosome derived Xist transcripts are first seen in 4-cell stage of female embryos prior to X-inactivation15,16. Specific inactivation of the paternal X-chromosome in the earliest stages of embryos correlates with the specific patterns of imprinting of paternal and maternal X-chromosomes; during spermatogenesis, the Xist locus is demethylated, passed on to the zygote in a hypomethylated state and, therefore, poised for transcriptional activity while the maternally derived Xist allele is fully methylated at this stage of embryonic development17. At some unspecified later stage of embryonic development in eutherians, the paternal imprinting is lost and Xist gene of one of the two homologues is demethylated randomly and this sets the stage for random inactivation of one of the two X-chromosomes18. It is notable that during meiosis in male mammals, the X-chromosome is inactivated and so this tissue is Xist gene active. Therefore, Xist transcripts are believed to be responsible for inactivation of the X-chromosome during spermatogenesis in a manner analogous to the X-inactivation in somatic cells of females19,20.

The mechanism of action of Xist transcripts in initiating inactivation of the X-chromosome from which these are produced is not known. Although Brown et al.21 showed by in situ hybridization that in interphase nuclei from female, the Xist transcripts were seen more abundantly in vicinity of the Barr body, it was not clear if this indicated binding of these transcripts to the inactive X-chromosome or to the nascent transcripts made by this chromosome. A structural feature of the inactive X-chromosome is worth noting: in interphase nuclei the two telomeric regions of the inactive X-chromosome (Barr body) remain closer22; the inactive X-chromosome in metaphase cells also shows a characteristic bend at the Xc locus23 and in the primary spermatocytes, the inactive X-chromosome shows a similar spatial orientation within the sex-plate24. This bending at the Xc may facilitate non-homologous chromatin association leading to heterochromatinization involving heterochromatin-specific proteins25. Whether this change in chromosome structure is due to the act of transcription at this locus or due to binding of the Xist transcripts or due to some other factors recruited by the Xist transcripts remains unknown26,27,28. Buda et al.29, using quantitative RT-PCR single nucleotide primer expansion assay found only about 2000 Xist transcripts per cell and suggested that only models that do not require Xist RNA to cover the entire inactive X-chromosome are compatible with the number of these transcripts present in a nucleus.

H19

This is another well-known example of a non-coding gene in mammals. The H19 gene was first identified as a CDNA that was coordinately regulated along with the α-fetoprotein by trans-actingraf in murine fetal liver30. Subsequent cloning of the human H19 homologue31 and comparison of sequence of the murine and human homologues revealed a lack of conservation of the small open reading frames although the organization of exons and introns was conserved and the base sequence of certain other regions was conserved32. These features led to the inference that H19 transcripts are not translated but as such function as RNA. In Southern blots, an H19 cognate could be detected in monkey, rat, and chicken but not in Droso phila33. A large proportion of H19 RNA in both human and mouse cells exists in association with 28S cytoplasmic particles34. Subsequent studies showed the H19 gene to be imprinted with only the maternal allele normally expressed; H19 expression has an interesting regulatory effect on
expression of the adjacent group of imprinted genes. The insulin-like growth factor 2 (IGF2) gene is immediately upstream of the H19 on human chromosome 7 and is expressed only from the paternal allele due to imprinting. A variety of studies have shown that H19 expression is specifically responsible for silencing of the neighboring six-located genes since the paternal H19 is imprinted (methylated); this promoter is transcriptionally inactive while the non-methylated maternal allele is transcribed. This in turn inhibits the cis-located IGF2 and other adjacent genes on the maternal chromosome. These neighboring genes are transcribed from the paternal chromosome on which the H19 allele is inactive. In this respect, H19 functions in a manner reminiscent of XIST, while XIST activity inactivates a whole X-chromosome, H19 expression influences a nearby imprinted domain.

The H19 is expressed as abundant, spliced and poly(A)-containing transcripts whose levels increase with cellular differentiation but are absent or reduced in several tumors. In agreement with these observations, Hao et al.11 have shown that H19 RNA has a tumor-suppressor activity. H19 RNA has also been suggested to have important roles in differentiation of embryonic cytodifferentiation. Thus, H19 RNA which does not have any appreciable open-reading frame.

An interesting feature of the XIST and H19 genes, shared with several other non-coding genes (e.g. the 93D locus of Drosophila), is the rather high degree of divergence of the base sequence (particularly at the open-reading frames) in spite of the structural organization of the locus being conserved. While in the case of human and mouse H19 genes, the exon-intron organization is highly conserved, in the case of XIST and Yar genes, the number of exons varies, although there is some degree of similarity. The XIST and H19 transcripts share similar secondary structures with long energetically favorable stem-loop structures; in both cases, the length of the stem-loops are present in the regions that show most conserved base sequence and therefore, appear to be functionally important. The mouse and human XIST/ YIST genes share certain short tandem repeats throughout their length at computable positions although their total numbers vary between the two species. H19 transcripts do not have such extensive repeat motifs but both the human and mouse H19 transcripts carry 8-10 copies of TGGGCG motif in a short region near the 3' end. Conservation of these repeat motifs is reminiscent of the conservation of a 9 bp motif in the otherwise divergent repeat units at the 3' end of the 93D locus of Drosophila. In all probability such short repeat motifs help in functions of these transcripts perhaps by determining some aspect of the secondary structure of the RNA and/or by providing binding sites for other molecules.

Other examples of non-coding genes in mammals

While the above examples of non-coding transcripts are better known, more cases of non-coding transcripts have been reported from mammals and other organisms in recent years. Some examples of these are briefly considered below.

H19-1 locus in mouse

The H19-1 locus in mouse has been reported to be a exon mutation site for retroviral insertions leading to myeloid leukemias. Insertion leads to alteration of the H19-1 locus which produces a 3 kb RNA derived from a gene consisting of 3 exons spanning 6 kb on mouse chromosome 2. This gene is conserved as a single copy gene in vertebrates and Drosophila; in mouse it is highly active in transformed myeloid cells but not in the normal cells examined and produces spliced poly(A) RNA which does not have any appreciable open-reading frame.

Synapse-associated non-coding RNA in rat

Veileca et al.12 identified a novel synapse-associated RNA, the H1-4 RNA, in rat diaphragm muscle; this transcript is present selectively in association with synapses in the synaptic zone of skeletal muscle of rat diaphragm and is upregulated during early postnatal development and after denervation. The H1-4 gene is without introns, yet produces 2 different sized transcripts with identical polyadenylated 3' ends. Sequence analysis revealed absence of any significant open-reading frames and, therefore, believed to function through its RNA products.

Human UHG for U22 snRNA

The nucleotides associated small RNAs (snorNAs) are involved in maturation of the IS ribosomal RNA and are usually produced by processing of intron fragments of protein-coding host genes. The U22 snORNA (earlier called human RNA Y910) is highly conserved between man and Xenopus. A search for its host gene in humans (the UHG) whose intron is processed to produce the U22 snORNA, revealed that the host gene specifies a poly(A) but non-protein coding RNA.
IPW or the imprinted gene in the Prader-Willi syndrome region was identified to map to the smallest deletion overlap corresponding to the Prader-Willi syndrome region on the proximal human chromosome 15q. Only the paternal allele is expressed in lymphoblasts and fibroblasts due to imprinting; the transcript is spliced and polyadenylated but its longest open-reading frame codes only for 45 amino acids. Therefore, this gene also appears to be making a non-protein-coding transcript. It has been suggested that the Prader-Willi syndrome phenotype may be a direct consequence of lack of expression of this gene.

Non-translatable repetitive transcripts in eggs Non-translatable RNA transcripts carrying single copy genomic sequences interspersed with repetitive sequences are commonly present in the cytoplasmic poly(A) RNA fractions in oocytes of sea urchins as well as Xenopus. These interspersed (Sp) maternal RNAs show a high sequence complexity, are distinct from nuclear as well as embryonic messenger RNAs and are not associated with polyadenylates. These Sp RNAs are not due to readthrough transcription of oocyte-specific mRNAs promoters like the histone genes but are specific transcripts in their own right; the presence of identical transcripts in somatic cells of late embryos also suggests that these transcripts do not result from readthrough products characteristic of oocytes.

A very interesting class of repetitive transcripts in Xenopus oocytes is the Xhrts (Xenopus hrts). Short interspersed repeat transcripts; these are a family of interspersed repeat RNAs that carry 3 to 13 repeat units of 79 to 81 bp length flanked by unique sequences. The repeat units carry a sequence that is very similar to the consensus repeat found in the human XMT1 and mouse Atu genes. The Xhrts transcripts show a characteristic in situ localization in growing oocytes; particularly remarkable is their overlapping distribution at the vegetal cortex in stage 3-4 oocytes with the Vgl and Xcr1 mRNAs (the Vgl is a GFP-like molecule while the Xcr1 is a NANOS-like molecule and both are implicated in axial patterning of early amphibian embryos). Roel and Ebitani showed that destruction of Xhrts in growing oocytes by microinjected antisense oligonucleotides resulted in delocalization of the Vgl, but not of Xcr1 RNA, suggesting that the Xhrts help anchor the Vgl transcripts on the microtubules at the vegetal pole in oocytes.

miRNA in fission yeast Nutrient starvation of diploid fission yeast (Sclerotia cerebrum) parasitic cells triggers them to enter meiosis through a cascade of events initiated by lowering of cAMP levels. One of cAMP regulated genes which is crucial for progression of meiosis in these cells is the mei2 gene which encodes an RNA-binding protein, mei2* function is required not only for pre-meiotic DNA replication but also for entry into meiosis I. Watanabe and Yamamoto showed that the RNA-binding mei2 protein interacts with 440 and 508 nucleotides mei2 mRNA produced by the smt2 gene. Neither of these two transcripts have any appreciable open-reading frames. This interaction was necessary for entry of cells into meiosis I but not for initiation of pre-meiotic DNA synthesis for which a different RNA may associate with the mei2 protein.

Non-coding genes: selfish DNA or genes with important biological roles? The belief that a spliced and polya containing RNA must have a protein coding function is so deeply entrenched that for any newly discovered instance where the transcript does not appear to code for a protein, the authors tend to make an apologetic explanation that the RNA may code for a protein under some unknown kinds of editing or alternative splicing events or else the RNA may actually be a product of a selfish DNA. Fortunately, the increasing number of such genes being known in diverse organisms has lent credence to the concept of non-coding transcripts also having a biological role. The non-coding transcripts may function as ribosymes or as antisense RNA regulating the activity of other transcripts as in the case of the lin-4 gene of C. elegans.
ological roles through RNA-binding. Contrary to earlier beliefs, it is now clear that in the ribosomes also, it is the RNA rather than the protein moiety that has the major catalytic activity. Thus, it remains possible that in the case of the various RNA polymersase or ribonucleoprotein modules, binding of RNA polymerase II transcribed non-coding RNAs to proteins may either alter the activity of the protein or may cause the RNA to have some activity about which we still do not know much.

A large number of coding transcripts have 3′ untranslated regions (3′UTR) of varying length; in a few cases these 3′ UTRs may be even longer than the coding region. Recent studies have shown that the 3′ UTRs play very important roles in either targeting the transcript to specific cell compartments or in controlling the kinetics of turnover of the mRNA or even in transcriptional activation of other genes as trans-acting factors. The 3′ UTR of alpha-tropomyosin mRNA can suppress tumourigenicity. In analogy with the 3′ UTRs, it is possible that the non-coding poly(A) containing transcripts may also carry out a variety of functions in the cell through their protein-binding properties. Certain zinc-finger proteins bind specifically to RNA-RNA hybrids with implications of their biological roles. In many of these cases the binding is dependent upon the secondary structure of RNA rather than its primary base sequence. Structural studies of the non-coding transcripts show conserved secondary structure although not necessarily the primary base sequence. It is likely that these RNA-RNA interactions function to regulate their activity by binding to the major groove of double helix. Sequences-specific binding of RNA or ribonucleoprotein to duplex DNA has also been considered to be important in gene regulation.

It is obvious that RNA-polymersase II dependent non-coding transcripts are no longer more enigmatic. They are now known to have physiological roles. These seem to have established themselves as a distinct class of genes with very important functions. Understanding the significance of such genes has been described in the common "selfish genetic element" label applied to them. Recent years have witnessed an increasing understanding of the biological significance of heterochromatin in view of which the biological relevance of the so-called "selfish" or "junk" DNA is inevitable. It is to be hoped that with an increasing awareness of genomic functions through structural motifs as well as through their primary base sequence alone, will stimulate an appreciation and understanding of this interesting class of genes. With RNA being the first "living molecule," it is but to be expected that even today biological systems continue to utilize this versatile molecule directly. Progress in modeling the structure of RNA and RNP molecules will be of considerable help in this direction.

Acknowledgement

Work in my laboratory on the 93K heat shock locus of Drosophila has been supported by the Department of Science and Technology, Govt. of India (New Delhi) and by the Department of Atomic Energy, Govt. of India (Bombay). I thank Dr. Rajiva Ramai for critical comments on the manuscript.

References


