RNA polymerase II dependent genes that do not code for protein

S.C. Lakshmi
Cytogenetics Laboratory, Department of Zoology, Banaras Hindu University, Varanasi 221 005
Received 14 August 1995; accepted 21 December 1995

In recent years more and more examples of RNA polymerase II dependent non-coding transcripts have been described. Although these have frequently been ignored as "selfish DNA elements", 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 considered. 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 secondary structures, rather than the primary base sequence of these transcripts. While some of these transcripts may function as ribozymes or as auto-regulatory, most others may function more directly through their specific protein-binding properties. Since RNA is believed to be the first "living" molecule, 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 polymerases, each earmarked for transcribing specific classes 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 5S 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 translocating the transcripts made by the RNA polymerase II which transcribes all the protein-coding genes dispersed throughout the chromosomal DNA. The RNA polymerase II transcribed genes share many features in common with regard to their transcriptional regulation, post-transcriptional processing of the primary transcript (heterogeneous nuclear RNA or hnRNA) and finally the transport of the processed messenger RNA (mRNA) to the cytoplasm for translation into their specific polypeptide products. The central dogma of molecular biology has described such an important role to the translational 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, while the whole of the Y-chromosome is condensed, heterochromatic and genetically "silent" and, therefore, its essential role in male fertility remained enigmatic although conventional genetic studies did identify a number of Y-linked "fertility genes" having specific effects on spermatogenesis. Meyer and his group14 showed that during the primary spermatocyte stage in D. hydei, the Y-chromosome opens up into very large microscopically visible and transcriptionally very active "flambrush" loops which are essential for normal differentiation of spermatids into mature motile spermatozoa. In D. hydei, six distinctive loops with characteristic morphology and transcript patterns are present while in D. melanogaster, the Y-chromosome loops in primary spermatocytes 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-chromosome linked fertility genes and the lambrush chromomeres 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 4000 kb in D. melanogaster and between 260 to 1500 kb in D. hydei18. Bulk of these huge transcription units is comprised of simple, satellite and of functional or degenerating transposable elements. None of these have any substantial open reading frames19. 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 intermediate chain16. The kis complementation unit of the Y-chromosome, at which the dynin gene maps, is estimated to be about 1200 kb long while the dynin transcript is estimated to be only about 14 kb19,21,22. 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 kis transcription unit16 or may be scrambled in its 1300 kb region23. All the known Y-chromosomal transcription units, active only during the primary spermatocyte stage, share a few common epigenetic features. This includes a) production of very large sized transcripts that remain restricted to the nucleus, b) the transcription units comprise essentially of simple repetitive and transposable elements which show remarkable sequence diversity between related species, c) the basic 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 bind to specific proteins and are responsible for the characteristic shape and size of the different Y-chromosomal lambrush loops15,16,18. It appears that during the evolution of these male fertility genes, many retrotransposons got incorporated within the lambrush transcription units. Interestingly such insertions do not appear to have affected the functioning of these genes. Thus the male fertility gene Q in the lambrush loops called "Noose" 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 loop16. Transcription of these retroposons is required but not sufficient for the function of fertility genes like the Q gene (Nooses) of D. hydei16,18. It is also interesting to note that the many tandem repeats of retroposons that are present in the loop transcription units show identical sense orientation although outside the loop boundary the same retroposons show random arrangement24. What is the function of these unusually large and apparently untranslatable RNA5? The Y-linked fertility genes and the lambrush loops are essential for spermiongenesis to proceed normally. The discovery of at least one protein-coding gene (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 large transcription units comprising of satellite sequences and transposons, have evolved because the fertility genes are "the ultimate heavens for selfish gene elements"16. 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-chromosome loops, a few other non-coding "genes" active during spermiongenesis in Drosophila are also noteworthy. One of the them is the enigmatic "cristal" or "Suppressor of sterile-Small" set of genes. D. melanogaster males with a Y-chromosome [X/O males] show presence of needle-like or star-like inclusions in their primary spermatocytes, a high level of non-dispersions of mitotic chromosome, abnormal distribution of organelles in meiotic and events of meiotic drive26. A Y-chromosome linked locus, the "crystal" or "cry" (also named "Suppressor of sterile-Small", Sst1,Sst2) and an X-linked locus, the "Small" or Sst2, are involved in this phenotype of X/O males D. melanogaster flies. These two X- and Y-linked loci contain arrays of partially homologous, randomly repeated sequences27. The normal functioning of cry or Sst2 locus requires a critical number of subunits rather than physical integrity of the whole array of tandem repeats27,28. The Sst allele of Small locus carries about 200 copies of the repeat unit while the Sts allele has only about 20 copies of the repeat units; in the absence of Sst2 gene on the Y-chromosome, the Sst allele causes production of needle-like inclusions in
xeromatoses while the Ste allele causes star-like
inclusions\textsuperscript{12}. The Ste\textsubscript{locus} codes for a pro-
tein which shares homology with the beta subunit of
casein kinase II\textsuperscript{12}. The Ste\textsubscript{locus} located on the
Y-chromosome suppresses activity of the Ste\textsubscript{locus} on the
X-chromosome so that in XY indi-
viduals, only a few stellate transcripts are made
and even those are not properly spliced and pro-
cessed but in the absence of Y-chromosome or in
the presence of Y-chromosome deficient for the
Ste\textsubscript{locus}, the level of stellate transcripts in
xeromatoses is significantly elevated resulting in
the formation of nucleoli on the star-like inclusions
and to consequently affect disruption and seque-
tration of homologous chromosomes and cytoplasmic
organelles like mitochondria\textsuperscript{2,9}. The Su(Ste) locus besides simulating homology with the Ste
locus, also shares sequence homology with the Her\textsubscript{t}
family of transformones\textsuperscript{3}. The 2800 bp repeat unit
of the Su(Ste) locus has been shown to consist of
(i) a region of homology to the Ste locus, (ii) a
Y-specific \textit{Ad} rich segment and (iii) a mobile ele-
ment 1650 inserted in the Ste sequence\textsuperscript{11}. It is not
known if the Su(Ste) transcripts code for a pro-
tein but this appears unlikely in view of its struc-
ture. The suppression of Ste activity in the pres-
ence of Su(Ste) has been suggested\textsuperscript{2} to be due to a
mutual competition for a limiting set of trans-
scription and splicing elements so that in the abse-
nce of Ste, these will be available to Ste for an
abundant production of the stellate protein. This
mutual interaction of these two genes has also
been considered an intrinsically type of selfish
genetic system\textsuperscript{2} since related species of \textit{Drosoph-
lla} do not have Ste or Su(Ste) loci. However,
the important point to note is that in this case al-
so, the Su(Ste) gene has very vital effect on
the organism without possibly having a typical protein
product.
Another gene active in the male germ line of \textit{D. metanagonaster}, but apparently not having a protein
product, is the recently identified\textsuperscript{8} Mrt 40 iso-
sequence (\textit{Male specific transcript} 40) located on the
section 40 of the right arm of chromosome 2: the
Mrt 40 sequence is organized as tandemly arrayed
1.4 kb repeat unit with transcripts limited to muc-
el 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 \textit{D. metanagonaster} tested although other species of \textit{Drosophila} did not show its presence\textsuperscript{2}. A function for this gene is
yet to be found.
Among the diverse variety of transcripts that
make up the bulk of Y-chromosome in \textit{Drosophila-
lla} is the \textit{microspia} family. These retrotransposons are
present both on the autosomes, X-chromosome
and the Y-chromosome of \textit{D. yakuba} as well as \textit{D. melanogaster}. Lankenau et al\textsuperscript{30} have reported
that in meiotic cells of male \textit{D. yakuba}, in addition to
full length transcripts, \textit{microspia} also encodes antisense transcripts complementary to the re-
verse transcriptase and RNAase H coding region; these antisense transcripts are present only in
male germ line since they are produced from a
testerspecific promoter. Furthermore, while most
of the sense transcripts are present as part of
giant \textit{RNA} molecules because of their location on
the longloop forming sites, the antisense \textit{microspia} transcripts are 1.0 and 1.6 kb long\textsuperscript{30}. It
appears that these full length antisense transcripts have an important role in the control of \textit{microspia}
encoded reverse transcriptase protein in male germ line\textsuperscript{24}.

The 93D or the HS\textsubscript{R} gene of \textit{Drosophila}
One of the first new protein coding genes to be
characterized in some detail is the 93D or the
\textit{hara} gene of \textit{D. reagonaster} and its homologues
in other species of \textit{Drosophila}. Transcriptionally, this is one of the most active genes following heat
shock\textsuperscript{23,24}. That this gene was different from other
heat shock genes was revealed by its singular in-
distinguishability in polytene cells treated with a variety
of agents like benznidazole, colchicine etc. (see re-
views in refs. 32-34). Since no new proteins were
induced when this locus was selectively activated by benznidazole, Laktash and Yukinose\textsuperscript{8} sug-
gested that this gene does not code for any protein;
this was confirmed when this gene and its iso-
logues in other species were cloned and se-
quenced\textsuperscript{23,25,26}. Sequence analysis revealed a remark-
ably conserved overall organisation of this gene
but equally remarkable divergence in the gene se-
quence in different \textit{Drosophila} species. In all spe-
cies, the locus spans more than 10 kb and in-
cludes 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 units\textsuperscript{25,26}. Although the
repeat units do not share homology between dif-
f erent species, all the repeats in a tandem array in
a species are highly homologous and maintain a
certain minimum and maximum length\textsuperscript{25}. The
\textit{hara} locus produces two primary nuclear tran-
scripts of >10 kb and ~1.9 kb size respectively;
the 1.2 kb transcript spanning the two exons and one intron is spliced to produce the 3.2 kb poly(A) [sic] [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 species. The relative abundance of the three transcripts predicted by this locus depends upon the nature of induction. Unlike most genes, the spliced out intron in this case is highly stable. Besides being induced by the different inducers, this locus is also developmentally active in most tissue types of embryo, larva, pupa and adult and although without a protein coding function, it is essential for survival of flies and also for development of normal threedimensional larva. Hsp 63 is known to bind to the heat shock induced 62D locus and recent observations in our laboratory suggest that the lethality due to deficiency of this locus is enhanced in 61S IS sex mutant heterozygotes.

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 aif transcripts. The rates of synthesis and/or turnover from the site of synthesis of the 87A-type and 87C-type hsp 70 and the aif transcripts varied in relation to the specific profile of the 93D transcripts present or synthesized in response to a given condition of heat shock. 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 3'untranslated regions (3' UTR) show considerable divergence. Shmema and Lakshmi suggested that the differing 7 UTR's may target the hsp 70 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 stress. 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 hsp 70.

The aif 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 locus, site for 3 copies of hsp 70 genes, also harbours about 10-14 copies of aif repeats. In addition, the aif repeats are also present at the heterochromatic chromocentre but these are not heat inducible. Those at the 87C site are heat inducible due to their being associated with sequence elements (the y elements) that are identical to the hsp 70 promoter region. The aif repeats that are immediately downstream of the the y 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 sizes. None of these appear to code for any protein. The suggestion that these sequences are one 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 aif repeats at the 87C or at any other site. 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 aif sequences during heat shock. Colchicine treatment which attunes transcription at the 93D locus (see above) also leads to an increase in the level of aif transcripts at the 87C site. 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 encoding for any protein have been identified in different mammalian genomes. These are briefly considered below.

The X1a 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 females. While the paternal X chromosome is preferentially inactivated in marsupials, the inactivation of one of the two X chromosomes in eutherians 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 translocation, 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 initiation of inactivation but also for its spread to
the entire chromosome length (reviews in refs 58–
60). A major 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 Xact gene. This gene, termed XIST (human) or Max (mouse) is X 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 17 kb long and the mouse XIST mRNA is 15 kb long but none of these appears to code for any protein 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 interphase nuclei. 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 and primitive endoderm lineages with exclusive inactivation of the paternal derived X-chromosome in all cells. X-inactivation in embryonic lineages occurs later and this is in contrast with respect to the parental origin of the X-chromosome. Paternal X-chromosome derived XIST transcripts are first seen in cell-stage female embryos prior to X (acetate/propionate). 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 egg in a hypothymelated state and; therefore, poised for transcriptional activity while the maternally derived Xist allele is fully methylated at this stage of embryonic development. At a somewhat 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-chromosomes. It is notable that during meiosis in male mammals, the X-chromosome is inactivated and so little of its XIST gene is 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 females.

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 showed by in situ hybridization that in interphase nuclei from female, the XIST transcripts are 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 if 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 closer; the inactive X-chromosome in metaphase cells also shows a characteristic bend at the Xic locus and in the primary spermatocytes, the inactive X-chromosome shows a similar spatial orientation within the sex-cord. This bending at the Xic may facilitate non-homologous chromosomal association leading to heterochromatinization involving heterochromatin-specific proteins.

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 unknown. Bustin also showed using quantitative RT-PCR single nucleoside primer extension 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 x-lysogenic proteins by trans-acting FL in mouse fetal liver. Subsequent cloning of the human H19 homologue and comparison of sequence of the murine and human homologues revealed lack of conservation of the small open reading frames although the organization of exons and introns was comparable and the base sequence of certain other regions was conserved. These features led to the inference that H19 transcripts are not translated but as such function as RNA. In Southern blots, an FL probe could be detected in monkey, rat, and chicken but not in Drosophila. A large proportion of H19 RNA in both human and mice cells exists in association with 28S cytoplasmic particles. 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 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 (methylation); this promoter is transcriptionally active 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 suppresses 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 some tumors. In agreement with these observations, Hu et al. have shown that H19 RNA has a tumor-suppressor activity. H19 RNA has also been suggested to have important roles in differentiation of embryonic cytотrophoblasts of cervical origin.

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 loci 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 Xnr 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 longest 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 H19/VNTR genes share certain short tandem repeats throughout their length at comparable 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 TGGCGG 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 locus in mouse

The H19 locus in mouse has been reported to be a candidate site for retrieving insertions leading to myeloid leukemias. Viral insertion leads to activation of the H19 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

Veleta et al. identified a novel synapse-associated RNA, the H14 RNA, in rat diaphragm muscle; this transcript is present selectively in association with synapses in the myofibular zone of skeletal muscle of rat diaphragm and is upregulated during early postnatal development and after denervation. The H14 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 is therefore, believed to function through its RNA products.

Human UHG for U22 snoRNA

The nucleotides associated small RNAs (snoRNA) are involved in maturation of the 18S ribosomal RNA and are usually produced by processing of intron fragments of protein-coding host genes. The U22 snoRNA (earlier called human RNA Y9-Y10) 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.
melRNA in fission yeast

Nutritional starvation of diploid fission yeast (Schizosaccharomyces pombe) cells triggers them to enter meiosis through a cascade of events initiated by lowering of CAMP levels. One of CAMP regulated gene which is crucial for progression of meiosis in these cells is the mei2Δ gene which encodes an RNA-binding protein, mei2Δ. 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 mei2Δ protein interacts with 440 and 508 nucleotides long mei2RNA produced by the mei2Δ gene. Neither of these two transcripts have any appreciable open-reading frame. 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 mei2Δ protein. It is notable in this context that meiosis-specific small nuclear RNA and a group of poly(A) RNA (aggRNA) has been described in fission yeast. The meiRNA appears to be different from the meiosis-specific RNAs of Drosophila although it is possible that the sgoRNA of Drosophila encodes a meiRNA counterpart. Whether the mei2Δ-melRNA complex may act as a meiosis-specific splicing factor. Whether this turns out to be the case remains to be seen but it is obvious that a non-coding RNA has a direct role in regulating sexual development in fission yeast.

Non-coding genes: selfish DNA or genes with important biological roles?

The belief that a spliced and poly(A) containing RNA must have a protein coding function is too 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 apodogisic explanation that the RNA may code for a protein after 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 ribozymes or antisense RNA regulating the activity of other transcripts as in the case of lin-4 gene of Caenorhabditis elegans.

In addition to roles as ribozymes or as antisense regulators, it is likely that the non-coding transcripts have more direct roles in cell regulation. Several possibilities exist. A large number of proteins are now known to perform significant bi-
ological roles through RNA-binding. 10-14. Contrary to earlier beliefs, it is now clear that the RNA not only is the RNA rather than the protein moiety that has the major catalytic activity. 15-19. Thus it remains possible that in the extremely varied RNA polymerase II transcripted 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 region (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 have very important roles in either targeting the transcript to specific cell compartments, 11,12,13, or in controlling the kinetics of turnover of the mRNA. 14,15 or even in transcriptional activation of cases or other genes as transcription factors. 16-18. The 3'UTR of alpha-1-antitrypsin of mice can suppress tumorigenicity. 19. 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 DNA
RNA hybrids with implications of their biological roles. 20,21. In many of these cases the binding is dependent upon the secondary structure of RNA rather than its primary base sequences. 22,23. Since most of the non-coding transcript segments conserved structure although not necessarily the primary base sequence. It is likely that these RNAs perform important biological functions through their structural motifs.

RNA directed de novo methylation of specific genomic sequences in plants has been demonstrated. 24. Various RNAs in crude HeLa cell extracts inhibit DNA methylase. 25. It is also known that certain unusual RNA structures favor cytosine methylation. 26. Thus it is possible that RNA-DNA hybrid may direct the DNA methylase. 27. It will be interesting to examine if the XIST RNA mediated inactivation of X-linked in mammal is due to similar mechanisms. The RNA-DNA triple helices are stable under physiological conditions as vitro 28,29 and it has been suggested that RNA may regulate gene activity by binding to the major groove of double helix. 30,31. Sequence-specific binding of RNA or ribonucleoprotein to duplex DNA has also been considered to be important in gene regulation. 32,33.

It is obvious that RNA polymerase II dependent non-coding transcripts are no longer mere curiosities of the biological variability. These seem to have established themselves as a distinct class of genes with very important functions. Understanding of the significance of such genes has been discussed in the common "selfish genetic element" label applied to them. Recent years have witnessed an increasing understanding of the biological significance 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 rather than through the 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 both to be expected that even today biological systems continue to utilize these versatile molecule directly. Progress in understanding the structure of RNA and RNP molecules will be of considerable help in this direction.

Acknowledgement

Work in my laboratory on the 93D 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 Ramann for critical comments on the manuscript.

References


100 INDIAN J. BIOCHEM. BIOPHYS. VOL. 33, APRIL 1996