[Major breakthrough in Nucleic Acid Research has not been sudden but rather slow. After the discovery of nucleic acid, it took a long time to realise its importance in cellular function. The involvement of DNA and RNA in heredity and the establishment of double helical structure of DNA in 1950s and the proposition of Central dogma, i.e., the information flows from DNA to RNA to protein along with decoding of information in DNA, accelerated the progress in nucleic acid research tremendously in the recent past. In this article Prof. Biswas has described the flow of message from DNA to RNA, i.e., transcription process mediated by RNA polymerase with which he has been associated from the very beginning. The controlling factors and recognition systems both in prokaryote and Eukaryote have been discussed. The implication of these recognition systems appear to play a vital role, and therefore, there are certain difficulties in expressing a eukaryotic gene to a prokaryotic system and vice versa as has been elucidated by recombinant DNA technology. —Ed.]

TRANSCRIPTIONAL CONTROL IN INFORMATION TRANSFER FROM DNA

B. B. BISWAS

Department of Biochemistry, Bose Institute, Calcutta 700 009, India

The information needed to carry out the cellular functions throughout the life span of a cell is stored in the genetic material known as DNA. Transcription is the first step in the flow of information from DNA to RNA, which in turn dictates the sequence of amino acids in proteins (translation), essential for cellular function. The process of transcription is mediated by the key enzyme RNA polymerase which catalyzes the synthesis of RNA by forming 3', 5'-phosphodiester bond using ribonucleoside triphosphates as substrates and DNA as template. This enzyme was discovered only two decades ago. Subsequent work showed many facets of the reaction. However, when initiation, elongation and termination of specific RNA chains were looked into, additional protein factors were found to be involved. The mode of action and precise control exerted by these factors on RNA synthesis has been studied and the overall process has been found to be more complex than it was envisaged previously. Several of events involved in the process of transcription by RNA polymerase, thus far elucidated can be summarized as follows: (i) the binding of RNA polymerase at specific sites on DNA template (initiation sites), (ii) the initiation of polymerisation (binding of first nucleotide to enzyme template complex), (iii) elongation of RNA chain from 5' to 3' end and (iv) termination and release of RNA chain. Synthesis of RNA chain begins generally with either adenine or guanine depending on the start signal (promoter) in the DNA template. From several lines of evidence, it appears that the additional initiation factor (σ) helps in the tight binding of
DNA at the start or initiation signal. The sequence of TATAATG (Pribnow box; anti-
sense strand), ten base pairs upstream of mRNA start point might play a role in recog-
nition of start signal for the transcription, though in this heptamer T at position 6 is
invariant in the different promoters so far sequenced. Thirtyfive base pairs upstream
sequences of the start point show also some homology as far as TTG adjacent to ACA
is concerned, though ACA is not well con-
served. Mutations in -10 and -35 regions affect the promoter function. Strong steric
constraints on the site at which RNA poly-
merase initiates transcription may determine the start point, since different starting positions
are used even though sequences are similar
adjacent to start site in different genes. Posi-
tive regulator (cyclic-AMP-receptor complex) in
certain operon recognizes a region located
some 55 to 70 base pairs upstream of start
site.

Elongation of RNA chain proceeds until
RNA polymerase reaches the stop signal which
like start signal consists of certain specific
nucleotide sequence such as AAATAAAA or
CAATCAA or its repeats, resulting in a run
of U at the 3' end of RNA or certain second-
dary structure (i.e., a stem and a loop struc-
ture). Termination mutant analyses also indi-
cate that the critical information for termina-
tion lies in the transcribed region of the
template and extends about 35 bases upstream
of the stop site. Translation process can also
affect in certain cases of transcription termina-
tion. Termination also depends in certain
cases on the presence of another factor (ρ fac-
tor). There are other accessory factors
discovered for the accurate transcription of
certain operon in vivo and in vitro. Thus it
is not only the primary sequence but also the
secondary and tertiary structure of DNA
which can control transcription.

In eukaryotes, there are at least three dis-
tinct types of RNA polymerases (in contrast
to single one in case of prokaryotes), the
first for messenger RNA, the second for ribo-

somal RNA and the third for transfer RNA
and 5S RNA synthesis. TATAAA (Hogness
box) similar to the Pribnow box of the pro-
karyotic promoter is found around position
-30 from the 1st nucleotide of the mRNA
in most eukaryotic genes. It now appears
clear that at least some of the sequences
required for the transcription of structural
genes by RNA polymerase II (mRNPase) are
localized in the DNA immediately flanking
the gene on the 5' side. This is however,
not the case with RNA polymerase III
(iRNPase) where these sequences can be
deleted without affecting transcription. In vivo
results point to a second DNA region loca-
lized further upstream, that is implicated in
transcription initiation. These also DNA
the template, is associated with proteins (both
basic and acidic) yielding repeated structures
defined as nucleosomes, which are finally
organized as the chromosomes. Each nucleo-
some consists of two each, of the four major
histones. The fifth histone (H5) is associated
with the linker region of the DNA between
such two nucleosomes. This H1 histone and
perhaps several nonhistone proteins result in
higher order of structures present in the
chromatin. As the information content
increases, say from E. coli to man by about
1,000 fold, the mere calculation of genes
varies from 5,000 in E. coli to 40,000 in man;
it appears then that most of the DNA in the
higher organisms consists not of the coding
regions, but of control regions. Thus multi
elemental control systems have been proposed
for transcription in eukaryotes. The sequences
coding for mRNA for globin or ovalbumin
are apparently located in different places in
the gene (see split gene concept). It seems
that the genes consist of informational
DNA (exons) interspaced with silent sequences
(introns). Thus transcription of these genes
entails only exon sequences with intron
sequences eliminated from the final transcript
or mRNA. So, similar to compaction of
genes, i.e., use of a single gene to code for
two different proteins in certain bacte
phages) and procaryotes, this split gene concept emerges from the studies with some animal viruses and eukaryotes. The question arises as to how the fragments of such messages are spliced or joined and what the intervening sequences (introns) are meant for. One thing is however, clear that the splicing is post transcriptional and the introns are eliminated. Thus the transcription unit in eukaryotes is larger than that in procaryotes. But it is at present perplexing as to what function is served by the genes remaining divided in eukaryote. RNA splicing activity may have a ribosome like structure involving a complex of structural proteins and RNAs with catalytic and specificity functions.

Another point that emerges is that the split message might help to produce variants of a single protein by differential splicing of the interrupted RNA. It might be specially applicable in the case of production of immunoglobulins. A relation between exons and protein functional units appears to have been established in lysozyme and haemoglobin; the central exonic region corresponds to a haem binding unit. There are two important differences between the genetic signals necessary for gene expression in procaryotes and eukaryotes (i) initiation signal for transcription and (ii) mRNA sequences at the 5' end necessary for translation into protein by ribosomes. A few base (3–12 bases) sequences, known as shine Dalgarno sequence (SD) occurs at the 5' end of procaryotic mRNA, is complementary to 3' end of 16S ribosomal RNA and this complementarity appears to play a role in stabilizing initiation complex between mRNA and ribosomes. mRNAs lacking this SD sequences are not efficiently translated by the E. coli cells. Thus manipulation by inserting right sequences upstream for a promoter for the proper expression of the desired gene is becoming a focal point in the case of an expression of eukaryotic genes in bacteria and vice versa. Further, from the structure of the chromosome, it emerges that the nucleic acid and protein interaction has a protective function leading to an altered structural configuration in the nucleic acid end ultimately to condensation so that most of the information in the nucleic acid is masked, allowing a small portion of the sequences to be transcribed as noticed during the different phases of cell growth. This rather justifies the statement that conformation is the information.