# Biology Today

## Filling in the Picture

Maura C. Flannery
*Department Editor*

Maura C. Flannery is Associate Professor of Biology at **St. John's University, Jamaica, NY 11439.** She earned a B.S. in biology from Marymount Manhattan College and an M.S., also in biology, from Boston College. Her major interest is in communicating science to the nonscientist. She has developed a biology course for criminal justice majors as well as courses in reproductive biology and in the future impact of biological research.

I often begin a class discussion of development with the statement: "The DNA in a fertilized egg contains the information necessary to produce everything in an adult individual." Aside from the obvious influence of the environment, this statement is essentially true, but it hides a great deal of ignorance. Despite all we know, development remains one of the great puzzles of biology. The question of how a linear DNA molecule can contain the information to form a vast array of three-dimensional structures is only now yielding to investigation. With recombinant DNA and other molecular techniques, we are beginning, but just beginning, to see how transcription of DNA sequences into RNA, and then translation of messenger RNA into protein, can produce the exquisite variety and complex beauty of living organisms.

The basic steps in transcription and translation were worked out for prokaryotes in the golden age of molecular biology in the 1950s and '60s. Then the major problem was figuring out the genetic code. At the same time, Jacob and Monod were working on their operon model of gene control in bacteria. Now researchers are studying transcription and translation in eukaryotes. A picture of beautifully orchestrated control is emerging, but it's still blurred. Recent work has presented scientists with a mass of facts about genetic control, but how all these pieces fit together is still uncertain. Though the final answers aren't even close, I think the excitement in this field makes it one that warrants the same kind of attention that I gave to gene organization and structure in last month's column.

The first problem to consider is how the cell knows which genes to transcribe. Evidence indicates that transcription is where the most crucial control is exerted during development, and the crucial step in transcription is the attachment of RNA polymerase to the proper DNA site for initiation. Ahead of, or "upstream" from, the coding sequence of a eukaryotic gene is a control region. Although many have investigated this area, Steven McKnight and Robert Kingsbury have developed a new technique which allows them to introduce clusters of point mutations into the control region without changing the length of this region (*Science*, July 23, 1982 and July 30, 1983). Previous mutational techniques produced additions or deletions; they changed the length of the region as well as its sequence, making results difficult to interpret because the spacing between DNA segments may be important in control.

McKnight and Kingsbury found three transcription control sequences within a region that extends for about 100 base pairs upstream from the first base encoding RNA. The control segment closest to the encoding region contains the "TATA box" which has already been identified as an important regulatory site. This sequence specifies the position where RNA synthesis is to begin and probably also specifies which DNA strand is to be transcribed; thus it controls the accuracy of transcription. The other two segments are further upstream. One region is rich in guanine, and the other is cytosine-rich. These two regions affect the efficiency of transcription initiation and therefore the frequency of transcription. These segments contain a six-base pair inverted repeat that could allow the two areas to pair with one another and form an intrastrand stem loop structure. This structure may be recognized by the RNA polymerase and aid its attachment to the control region. This mechanism is just speculation at the moment, but it is plausible, particularly since many lines of evidence indicate that the three-dimensional structure of the DNA is crucial to proper control. DNA is no longer seen as a mere linear sequence of nucleotides, but as a molecule with a three-dimensional structure as specific as that of a protein.

In simian virus 40 (SV40), there is another regulatory sequence fur-

ther upstream from the mRNA start point (*Nature,* May 6, 1982). Called the enhancer, it consists of a directly repeated 72 base pair sequence. Enhancers may regulate cell differentiation by acting as chromatin organizers; again, three-dimensional structure comes into play. There is evidence, also from SV40 research, that there are proteins that act as repressors of mRNA synthesis similar to those seen in prokaryotes. By binding to the DNA, they can suppress transcription. Thus, there may be many factors influencing initiation. They may not all affect the same genes, nor act at the same stage of development. Different circumstances call for different controls. Control of transcription termination is possible too. During adenovirus infection, the same transcription unit in the viral genome has different termination sites early and late in infection (*Nature,* June 3, 1982).

Control can also be exerted during RNA processing. In eukaryotes, the primary RNA transcript undergoes a series of changes that result in messenger RNA. These changes begin soon after initiation of the RNA chain. The addition of a cap, a methylated guanylate residue, to the 5′ end of the primary transcript seems to be an almost universal step in processing. Next, adenylic acid residues are added to create a 3′ poly(A) segment. The poly(A) is added at a point before the end of the transcription unit. An endonuclease cleaves the end of the primary RNA transcript before the poly(A) is attached. The enzyme cleaves at the right spot by recognizing the sequence AAUAAA, 10-25 bases upstream from the addition site. Recent results indicate that the function of poly(A) is to insure cytoplasmic stabilization of the mRNA. The poly(A) tail shortens with time; this may ultimately lead to the destruction of the messenger.

Not all nuclear RNA primary transcripts are fully processed. Though most of them are capped, only about a quarter of them have poly(A) tails. Whether some specific transcripts are processed more efficiently than others isn't known, but this may be another level of control. Specific tail-less transcripts may eventually be adenylated when they are needed as part of the stable mRNA pool in the cytoplasm.

Still another step in RNA processing is the splicing out of the introns—intervening sequences that do not code for amino acids. The mechanism of splicing isn't well understood; a particular enigma is how the introns can be snipped out so precisely. Heterogeneous ribonucleoprotein (hn-RNP) particles in the nucleus may be involved. Clusters of 10 to 15 of these particles may form a superstructure on which splicing occurs; again, three-dimensional structure seems important (*Science,* June 11, 1982).

The power of RNA in processing is illustrated by the precursor to ribosomal RNA in *Tetrahymena thermophila*. The RNA folds into a three-dimensional structure that specifies the sites of excision of an intron. The RNA also catalyzes the splicing. This autocatalytic activity of RNA is one of the recent surprises in molecular biology. If it turns out to be a widespread capacity, it will fundamentally change our concept of RNA.

The same primary transcript can be spliced in alternative ways leading to the production of two different proteins. For example, in thyroid cells, the RNA transcript from the calcitonin gene is processed to produce a messenger coding for the hormone calcitonin. In hypothalamic cells, the product of the same primary RNA transcript codes for a different protein called calcitonin gene-related peptide (CGRP). The beginning or 5′ ends of both messengers are identical, but beyond nucleotide 227, they encode completely different amino acid sequences (*Nature,* July 15, 1982). This is achieved by splicing a single 5′ sequence onto either a calcitonin-specific or a CGRP-specific encoding sequence or exon; the control of this differential splicing is tissue specific. The primary RNA transcripts made in the thyroid and in the hypothalamus are the same, thus transcription seems to proceed through both coding regions, no matter which mRNA is finally produced.

Alternative processing may be one way to generate a diversity of peptide hormones and increase the flexibility of endocrine control. In fact, alternative processing may be a relatively widespread phenomenon, going beyond the endocrine system (*Science,* January 1, 1982). Evidence from mouse genetics indicates that differential assembly of exons of a gene causes the formation of different variants of a given protein product at different stages of development.

How RNA is transported from the nucleus to the cytoplasm is another area of ignorance at the moment, so it's impossible to say if this is a control point. We don't know if some messengers are more efficiently transported than others. What is known is that once mRNAs reach the cytoplasm their half lives vary greatly within a single cell, and the half life of a single mRNA can be very different in different tissues. Differential destruction of mRNAs is obvious during erythroblast differentiation in mammals. The developing cell makes many mRNAs, with globin mRNA synthesis representing only a small fraction of the total. During the last four cell divisions, the cell becomes a highly specialized reticulocyte synthesizing over 90% globin. It appears that, while the globin mRNA is being actively translated, a very specific destruction mechanism eliminates most other mRNAs. How some mRNAs are stabilized

more effectively than others cannot yet be determined, but the rate of removal of the poly(A) sequence may be involved.

Though we haven't even touched on translation and the controls operating at the level, there are already a myriad of points during transcription and RNA processing where controls are exerted. Why so many? Multiple control mechanisms would seem to be essential for a process as complex as development; one or two control points couldn't be sufficiently effective. On the other hand, this doesn't mean that all these control points will turn out to be crucial. Most of these mechanisms have only been studied in one or two systems, and in many cases they were viral systems which may not be the best models for normal control processes. Some of these control points will prove to be important, but others will probably end up being of, at most, limited importance. Unfortunately, it's too early right now to discriminate, so all avenues of investigation are worth following. That's where the adventure of research lies. You can't always tell which lines of research are dead ends and which are yellow brick roads.

As far as translation is concerned, controls also seem to be exerted here because not all mRNA is translated at the same rate. This differential control is most obvious during early embryonic development. Protein synthesis increases at least 50-fold in a sea urchin egg when it is fertilized, even though the amount of mRNA present in the cytoplasm hasn't changed, and the proteins produced are similar (*Science,* July 2, 1982). Changes in the patterns of protein synthesis occur mainly during the transition from blastula to early gastrula, when differentiation becomes apparent in the embryo. In clam eggs, however, the rate of protein synthesis doesn't change after fertilization, but the proteins produced are different. Certain mRNAs, though present at all times, are specifically translated only before or after fertilization. Switches in protein production have been detected in nonembryonic cells as well. When mammalian cells grown in culture are subjected to heat shock, total protein synthesis decreases, but the translation of certain "heat-shock" proteins appears to increase.

The machinery of translation is complex. Many ribosomes, themselves intricate structures of rRNA and protein, are strung along a messenger to form a long polysome. Transfer RNAs are also necessary, as are the proper aminoacyl-tRNA synthetases to charge each specific tRNA with its specific amino acid. Additionally, protein translation factors are needed for a variety of functions, for example, attaching a charged tRNA to its acceptor site on the ribosome. Obviously, such a complex process is expensive for a cell, so it doesn't produce any more of these RNAs and proteins than it needs. Also, all the components must be manufactured in the proper proportions to avoid waste. Due to these considerations, the availability of the translation machinery could easily be a control point in protein synthesis. Though the eukaryotic system hasn't been thoroughly investigated, coordinated control of the synthesis of translation components does exist in prokaryotes. During exponential growth in *E. coli,* no more ribosomes are produced than are required for the maintenance of growth. There is also evidence for cotranscription of tRNA genes and genes for ribosomal proteins (*Nature,* December 3, 1982).

Once translation starts in eukaryotes it doesn't necessarily proceed uninterrupted. At least this is true for secretory proteins, where the first part of the growing polypeptide chain to emerge from the ribosomal complex is a "signal" sequence that directs the ribosomal complex to the endoplasmic reticulum (ER). A protein has recently been discovered that stops translation after the synthesis of the signal sequence by binding to the signal (*Nature,* June 24, 1982). This inhibition will only be released when the signal sequence makes contact with the ER, where a proteolytic enzyme releases the inhibiting protein. The nascent protein is then transferred across the ER bilayer. Stopping translation until a site on the ER has been found makes sense. If translation were to proceed unchecked, the polypeptide chain might be completed before it could be transferred to the membrane. The resulting protein would end up in the cytoplasm and never be able to fulfill its extracellular function.

Post-translational control is also possible. The same polypeptide can be broken down at different rates in different cells or at different times during development. Also, addition of various prosthetic groups can alter a protein's function. Finally, differential proteolysis can result in the production of different proteins from the same primary translation product; this appears to be how some pituitary cells produce β-lipotropin, while others produce adrenocorticotropin.

This report is necessarily sketchy. Since research in this area is just beginning, the available results don't yet form a coherent picture. It's like a paint-by-number picture in which only two or three colors have been filled in. Yet even at this early stage, the beauty of the picture is already obvious.