J. Indian Inst. Sci., Sept–Oct, 1991, **71**, 485–489.

Trends in molecular biology

G. PADMANABAN
Division of Biological Sciences, Indian Institute of Science, AND Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore 560 012, India.

Abstract

Molecular biology as a discipline was born with scepticism in the minds of at least some classical biologists and biochemists. However, it has truly evolved into a major discipline of modern biology and revolutionized our understanding of biology at the molecular level. The two fundamental questions, namely, the role of excess DNA in higher organisms and the structure and regulation of expression of eukaryotic genes are interrelated and a great deal of information is now available. It is proposed that at least part of the excess DNA is required to bring about the temporal and stage-specific regulation of gene expression. Knowledge at the molecular level has helped to go up the ladder and investigate the biology of the total organism. The fundamental concepts on gene structure and expression have led to the birth of new biotechnology with unlimited potential to solve problems of health and disease, food and agriculture and industry and environment.

Key words: Central dogma, C-value paradox, gene expression, future.

1. Protein and DNA paradigm

The birth of molecular biology signalled the change from protein to DNA paradigm in terms of understanding the nature of the gene. Even though Watson and Crick had propounded the double helical structure of DNA in 1953, textbooks in biology appearing even as late as 1957 and 1962 did not completely switch over to the DNA paradigm. For example, *Principles of modern biology* by Douglas Marsland (1957 edition) states: "Genetic material is DNP and that the specificity of gene comes from both the bases of DNA and the amino acid side chains of protein".

The term molecular biology did evoke strong sentiments. A talented scientist such as Edwin Chargaff continues to feel bitter and states that a molecular biologist is a piquant ragout and practises biochemistry without a licence! On the other hand, Hoagland states almost with poetic ecstasy: "while, we biochemical explorers were slashing our way through a dense jungle to discover a beautiful temple, Francis Crick, floating gracefully overhead on the gossamer wings of molecular biology,

* Based on a talk given to the faculty of the Jawaharlal Nehru Centre for Advanced Research, Bangalore, on November 23, 1990.

waited patiently for us to see the goal that he was already gazing down upon"[1]. Although the term molecular biology is all encompassing, its connotation has become clear. The pivotol message it seeks to convey rests on the central dogma:

$$\text{DNA} \rightleftharpoons \text{RNA} \longrightarrow \text{Protein.}$$

2. Molecular biology calendar[2]

The foundation for molecular biology was laid by Beadle and Tatum when they proposed in 1941 that a gene codes for an enzyme. A very brief calendar of events in the development of the field of molecular biology would read as follows:

1941 – Beadle and Tatum show that a gene codes for an enzyme.

1944 – Avery shows that DNA is the genetic material.

1953 – Waston and Crick propose the double helical structure for DNA.

1958 – Meselson and Stahl demonstrate the semi-conservative nature of replication.

1961 – Triplet nature of genetic code; concept of messenger RNA; Jacob and Monod's operon concept.

1970 – Temin and Baltimore identify the process of reverse transcription (RNA $\longleftarrow$ DNA).

1974 – Entry of recombinant DNA techniques.

1976 – Retroviral oncogenes identified as causative agents of cell transformation.

1977 – DNA sequencing becomes possible. Eukaryotic genes are split and have introns.

1979 – Cellular oncogenes discovered.

1981 – RNA has catalytic activity. Transgenics can facilitate study of gene regulation *in vivo*.

1987 – Polymerase chain reaction easily permits a million fold complication of any piece of DNA and therefore recombinant DNA manipulations become feasible at very low concentrations. A major initiative towards sequencing the human genome started.

3. C-value paradox[3]

The fundamental issues in molecular biology are basically two fold. One is that posed by C-value paradox, which highlights the enormous increase in DNA content with evolution, without a concomitant increase in the number of genes. For example, the increase in DNA content from Mycoplasma (10^6bp), through *E.coli* ($4{\cdot}6 \times 20^6$bp) and yeast ($2{\cdot}3 \times 10^7$bp) to birds, amphibians ($7{\cdot}8 \times 10^8$bp) and mammals (2×10^9bp) defies an answer in terms of the lack of a corresponding increase in the complement of genes. To an extent, the increase in DNA content is explained as due to the occurrence of non-coding flanking sequences, repetitive DNA elements and introns. But, why there is such an elaborate non-coding motif in which are

embedded the coding regions? At the same time, nature is not lavish in using the extra DNA available. In fact, there is conservation in the utilization of the available genetic information. It is known that only a fraction of the potential genes is turned on at any point of development of a higher organism. While house-keeping genes have to function all the time, many other genes are brought to play only at critical periods.

4. Control of gene expression[4,5]

This brings us to the second issue, namely, control of gene expression. The micro-environment of the cell, which is turn is influenced by the macro-environment, decides when and how a gene is turned on or off. The last decade has seen an explosive increase in our knowledge of the structure of eukaryotic genes. This became possible with the advent of recombinant DNA techniques. While genes could be identified at least on a functional basis, using mutations and genetics in prokaryotes, such an approach, barring a few examples, cannot work in eukaryotes. Recombinant DNA technology has enabled us to pick up a needle from a haystack—a single gene from the huge mass of DNA and amplify the same to obtain large quantities. The structure of the eukaryotic gene with its flanking sequences, introns and unique chromosomal locations, has turned in several surprises, when compared to the structure of a typical prokaryotic gene. The current excitement is in deciphering how this elaborate gene structure in the eukaryote responds to the environmental signals. Although gene expression is regulated at very many levels, a basic blueprint is available on the mechanism of transcriptional activation, which is fundamental to this process. The study of a large number of eukaryotic genes has revealed that activation or repression of a gene is brought about by interaction of protein factors with DNA sequences flanking the gene. Thus, these regulatory DNA sequences referred to as *cis*-acting elements and the proteins, referred to as *trans*-acting factors, interacting with the *cis*-acting elements, play a crucial role in the transcriptional activation process. Gene expression is also regulated by many other parameters such as chromatin structure, RNA stability, translation and post-translational modifications.

All this elaborate machinery of gene regulation leads to a situation where several proteins are required to regulate the production of one protein. In this, there are two categories. One category is the transcriptional, post-transcriptional and translational machinery common to all genes. The other category involves gene-specific protein factors. A major effort at the present time is to identify gene- and tissue-specific transcription factors and elucidate the mechanism of interaction with DNA leading either to activation or repression of the gene. This concept of gene regulation in eukaryotes is not very different from the principles of operon concept propounded by Jacob and Monod for prokaryotes, but substantially different in the matter of details and gene architecture. It also appears that a limited number of protein factors, perhaps in appropriate combinations, may be able to bring about the regulation of a large number of genes, but investing with gene and tissue specificities.

5. Is all the excess DNA useless?

An interesting consequence of all these studies is that the biochemical definition of a gene as coding for an enzyme or a protein has become blurred. The functional definition of a gene as the unit of heredity, irrespective of the structural complexity, is still easy to understand. Secondly, it has become difficult to identify the start and end regions of a gene in a functional sense. While it is easy to identify the coding region, namely, the exons and introns, involved in coding for the RNA and its subsequent processing, it has not been so easy to identify the *cis*-acting DNA sequences. There are examples where DNA sequences far away from the coding region regulate the expression of the gene. Although these regions are far away in terms of the number of bases, they are obviously close by to the site of initiation of transcription in three dimension. In fact, concepts of regulation based on protein-protein interaction, brought into juxtaposition as a result of DNA bending, elegantly explain the remote controls exerted on the expression of a gene[6].

Looking from this perspective it is rather difficult to dismiss all the excess DNA as useless or selfish DNA. It is true, as for example, that one is not able to explain why there should a 100-fold difference in DNA content among amphibians. At the same time, it is also clear that in eukaryotes a significant proportion of DNA could be involved in regulating the expression of a gene. This is not surprising, considering the fact that a gene has to respond or decide not to respond to myriads of environmental signals during the life cycle of an organism.

6. Looking to the future

The advent of recombinant DNA technique has ushered in a new phase in molecular biology. This phase concerns progress in terms of basic molecular biology and applications in biotechnology. The identification of gene defects in Duchenne muscular dystrophy and cystic fibrosis is an eloquent example of what this approach can do to elucidate molecular deformities in such intractable genetic disorders. Besides, the application of genetic engineering in terms of the production of useful proteins, preparation of safe vaccines, generation of transgenic plants, creation of engineered microorganisms for a variety of useful purposes, identification of genetic defects, etc., has given tremendous fillip to the field of biotechnology. From the perspective of basic studies, the present emphasis on understanding transcriptional controls of eukaryotic genes will continue. One also sees an attempt to understand the architecture and regulation of large gene assemblies at the chromosome level and eventually at the level of the whole organism. While a reductionist approach has been very essential to get into the molecular details, it does become important to examine how these details are coordinated in the whole organism. Such an approach is necessary to interpret growth, development and differentiation in molecular terms. That is why systems such as *C. elegans* and Arabidopsis have become popular. Besides, methodologies based on pulse-field electrophoresis and other manipulations to handle and analyse large chunks of DNA are in the process of further development. Mapping and sequencing of large genomes including the human genome would throw

light on the architecture of the DNA universe and the intended and unintended applications of this knowledge can be many. The stage is also perhaps set for understanding the molecular details of higher functions of the brain such as consciousness, memory, sleep and emotions. The gene basis of behaviour would be another exciting area. Despite the interest in all these esoteric questions, societal pressure has already led to intense activity in the area of molecular biology of AIDS virus. In a country like India, issues such as population control, eradication of infectious diseases, protection from environmental degradation and production of more food will and should dictate priorities in basic research and molecular biology has a major role to play in this regard.

References

1. *Biochem. Biophys. Acta*, 1989, **1000**, 1–482.
2. LEWIN, B. *Genes IV*, 1990, p. 803, Oxford University Press.
3. BRITTEN, R. J. AND DAVIDSON, E. H. Gene regulation for higher cells: A theory, *Scuence*, 1969, **165**, 349–357.
4. MANIATIS, T., GOODBOURN, S. AND FISCHER, J. A. Regulation of inducible and tissue-specific gene expression, *Science*, 1987, **236**, 1237–1245.
5. WASYLYK, B. Enhancers and transcription factors in the control of gene expression, *Biochem. Biophys. Acta*, 1988, **951**, 17–35.
6. PTASHNE, M. How eukaryotic transcriptional activators work, *Nature* (Lond.), 1988, **335**, 683–689.