The molecular basis of embryonic development

Edward J. Massaro

The information for the development of a complex adult organism from a relatively simple fertilized egg is contained within the deoxyribonucleic acid (genes) of the egg. Fundamentally, the genes are responsible for protein biosynthesis. This is accomplished through the mediation of ribonucleic acids (RNA). In the broadest terms, protein biosynthesis involves deoxyribonucleic acid (DNA) transcription into RNA and RNA translation into the linear amino acid sequence or primary structure of the polypeptide chain. Most organisms possess a large variety of proteins. These can be divided into two classes: (1) structural proteins and (2) enzymes. Enzymes control the nature and rates of the metabolic processes of cells. Not all of the metabolic processes characteristic of a given organism are found in every cell type of that organism. In fact, the enzyme complement of different cell types differs both qualitatively and quantitatively. Thus, the individuality of cells (i.e., their phenotype) is a reflection of the activities of their constituent enzymes. Since all cells of an organism possess the same genetic potential, this biochemical individuality is established at some time during embryonic development via the interaction of genome and environment. In essence, then, the basic principle of development in “higher” organisms is differential gene action. At present, little is known about the mechanism of regulation of gene action at the level of the gene. However, through studies of the end products of gene action, RNA, and protein, advances in our knowledge of the interrelationships of development and differential gene action have been made. For example, studies of the enzyme lactate dehydrogenase have been especially useful in providing insight into differential gene action during embryonic and phylogenetic development.

In recent years, the literature dealing with the central problem of embryonic development, the regulation of gene function, has grown enormously. This paper is not a definitive review of the literature. Its purpose is to present a general background to the work to be presented by my fellow participants in this symposium.

The potential of an organism to manifest a particular trait under a particular set of environmental conditions is transmitted from generation to generation. It is now well established that the hereditary information for the distinct morphological and physiological characteristics (the phenotype) of an organism is stored in and transmitted by the nucleic acids.\textsuperscript{1-3} Moreover, all hereditary characteristics are the end products of interacting biochemical processes and the hereditary determinants (the genotype) control the biosynthetic reactions of the cell within preset limits. The sum total of the interaction between the hereditary determinants and the environment results in the development of organisms with distinct phenotypes.
Fig. 1. The major pyrimidine and purine bases occurring in DNA and RNA.
In order to function as the repository of hereditary information, the chemical substances comprising the genetic determinants must possess certain fundamental attributes. They must be capable of being arranged into highly specific messages analogous to the arrangement of the letters of an alphabet into correctly punctuated sequences of words. Also, these chemical substances must be relatively stable and capable of exact replication in order to be transmitted without error to successive generations. Of the vast array of molecules comprising biological systems, only the nucleic acids possess the requirements necessary to function as informational molecules.

There are two types of nucleic acids of major biological importance, deoxyribonucleic acid (DNA) and ribonucleic acid (RNA). Both are long-chain polymers of nucleotides. The nucleotides are composed of a purine or pyrimidine base (Fig. 1) to which a phosphate and pentose sugar moiety are joined. In DNA, the sugar moiety is D-2-deoxyribose. In RNA, it is D-ribose. The most common pyrimidine bases in DNA are cytosine and thymine. In RNA, they are cytosine and uracil. Adenine and guanine are the most commonly occurring purine bases in DNA and RNA. The structural formulas of representative purine and pyrimidine nucleotides are shown in Fig. 2. The four principal nitrogenous bases function as the four letters of the genetic code and are abbreviated A, T, C, G. It is now known that the hereditary information is encoded in the sequence in which these four basic elements occur in the nucleotide polymers.

It is firmly established that all somatic cells of a complex organism receive identical chromosome sets by cell division. Each of these cells, then, potentially is capable of synthesizing the same types and quantities of molecules. However, it is obvious that the cells of higher organisms develop remarkably different characteristics. Thus, certain molecules appear in only one or a few kinds of cells. For example, except under highly abnormal conditions, the hemoglobin genes function only in certain cells of the erythropoietic series, insulin is produced exclusively in certain cells of the pancreas, and the pigment melanin is synthesized only in the melanocyte. Moreover, in metazoan cells a temporal dimension is evident; genes are activated at highly specific stages of cell differentiation. Once activated, they may continue to function for relatively long periods of time.

The fundamental problem of developmental genetics is the mechanism by
which gene function is regulated. For the most part, this problem remains unsolved at the level of the gene. However, in the final analysis, the individuality of cells (i.e., phenotype) is the end product of the activities of their constituent enzymes. These enzymes are a reflection of the makeup of that portion of the genome which is active in a particular cell type. The pathway from gene to protein is a multistep complex which is subject to many types of checks and balances. Fundamentally, the DNA of the active portions of the genome is transcribed into RNA which, in turn, is translated into the linear sequence of amino acids (primary structure) comprising proteins. By investigating the mechanism and control of protein synthesis it should be possible to gain insight into the nature of gene regulation at the molecular level.

Fig. 3 illustrates the basic steps in protein biosynthesis. DNA, messenger RNA (mRNA), soluble or transfer RNA (sRNA or tRNA), ribosomes, the cellular pool of free amino acids, and the enzymes, RNA polymerase, aminoacyl RNA synthetase, and peptide polymerase are all involved in the process. The genetic message encoded in the genome is transcribed from a single strand of the DNA double helix into mRNA.5-8 RNA polymerase catalyzes the assembly of ribonucleotides into mRNA.9-11 The messages dissociate from the genome and become associated with ribosomes in the cytoplasm to act as templates upon which polypeptide chains will be synthesized. Such clusters of mRNA and ribosomes are called polysomes. Amino acids are transported to the templates by sRNA molecules. The aminoacyl RNA synthetases select amino acids from the cellular pool of free amino acids. Selection is highly specific and only one kind of amino acid will react with a particular species of enzyme. The amino acids, then, are transferred to particular species of sRNA. Transfer is mediated by the same enzyme and is also highly specific. Through hydrogen bonding, the amino acid–sRNA complex pairs with a specific segment of the template mRNA. Amino acids are added one at a time and are linked via peptide bonds through the mediation of the enzyme, peptide polymerase. In this orderly manner, the polypeptide chain grows from N-terminal to C-terminal end.12,13 The growing polypeptide chain is bound to sRNA through its terminal carboxyl group. It also appears to be bound to the 50S ribosomal subunit;14 but this binding does not appear to require mRNA.

Ultimately, the polypeptide chains are completed and are released from the polysomes. Due to hydrogen bonding and the nature of the bond angles between adjacent amino acids, the polypeptide chains coil into their characteristic “secondary” configurations. Furthermore, as a result of various types of ionic and covalent interactions, these coiled chains fold into characteristic “tertiary” configurations. In many cases, individual polypeptide chains are nonfunctional until they have entered into aggregations of higher order (the “quaternary” structure of polymeric proteins).

The number, type, and sequence of amino acids comprising a polypeptide chain dictate its three-dimensional configuration.15 Thus, in the final analysis, the primary structure of a polypeptide chain is responsible for its physicochemical properties. Primary structure is directly controlled by the genetic determinants residing in the DNA of the cell. Through it, the genetic material controls the phenotypic potential of the cell. Maintenance of the vital functions of the cell depends on accurate translation of the information stored in its genetic material. Base pairing provides the mechanism by which this highly accurate translation is accomplished.

As shown in Fig. 4, the DNA molecule is composed of two complementary polynucleotide chains. These are joined together by hydrogen bonds between the nitrogenous bases of each chain, as the sides of a ladder are joined by the rungs, and the whole structure is twisted into the form of a double helix. Watson and

Fig. 4. The basic structure of the DNA molecule.
Crick hypothesized that the hydrogen bonding does not take place between random base pairs but, as a rule, adenine pairs with thymine and guanine pairs with cytosine (Fig. 5). DNA-RNA and RNA-RNA hybrid molecules are formed in accord with these same rules. In RNA, however, uracil replaces thymine. Thus, in DNA-RNA hybrids, adenine of DNA pairs with uracil of RNA. In the duplex structure formed by two RNA strands, adenine, likewise, pairs with uracil.

The first step in protein biosynthesis is the transfer of the information for polypeptide primary structure from the DNA of the genetic determinants to the ribosomes. This transfer is mediated by mRNA. A single strand of the DNA double helix functions as a template on which the RNA message is assembled. This involves the formation of a DNA-RNA duplex and, through the pairing of complementary bases, messenger RNA's are assembled which are precise complementary copies of...
Trinucleotide
(bases complementary to mRNA coding triplet)

Fig. 6. A model of the proposed structure of sRNA. (From Zubay, Geoffrey, Science 140: 1092, 1963. Reproduced by permission of Science, Washington, D. C.)

the DNA template. The completed messages dissociate from the template and enter the cytoplasm, where they bind to ribosomes. In these polyosomal associations, mRNA serves as a template for the orientation in a specific sequence of amino acid-charged sRNA molecules. This sequence is a reflection of the primary structure of a polypeptide chain. Thus, the genetic map and the sequence of amino acids in the polypeptide chains are co-linear. The highly specific alignment of the amino acid-charged sRNA molecules is dictated by the base sequence of the mRNA via the base pairing rules. Thus, the base sequence of the message selects the correct complementary base sequence from the pool of charged sRNA molecules.

Like mRNA, the sRNA molecules are assembled at specific regions of the genetic material. Therefore, each species of sRNA possesses a unique base sequence which is complementary to a particular region of the genetic material. The proposed structure of the sRNA molecule is shown in Fig. 6. It is a double helix formed from a single polynucleotide chain which is bent upon itself somewhere near the middle of the chain. Hydrogen bonds between complementary bases stabilize the structure. sRNA molecules contain two distinct binding sites—one for transporting amino acids and one for attachment to ribosomally bound RNA.

Apparently, all sRNA molecules have an identical terminal trinucleotide sequence, -cytidylic-cytidylic-adenylate. The amino acids are attached to the terminal adenyl acid residue through either the 2' or 3' cis-hydroxyl group of the sugar moiety (ref. Fig. 2). Thus, it would appear that all the amino acid binding sites are identical. However, it is firmly established that the interaction between sRNA species and amino acids is highly specific. This specificity is maintained by the aminoacyl RNA synthetases, which selectively bind specific amino acids from the cellular pool of free amino acids and catalyze amino acid activation: amino acid + adenosine triphosphate (ATP) + enzyme ⇌ aminoacyl – adenosine monophosphate (AMP) – enzyme + pyrophosphate (PP). The activated amino acids remain bound to their specific synthetases until they are transferred to specific sRNA molecules. The transfer reaction is catalyzed by the synthetases: aminoacyl-AMP-enzyme + sRNA ⇌ aminoacyl-sRNA + enzyme + AMP.

The mRNA-binding site of sRNA consists of a sequence of three nucleotides which is the exact Watson-Crick base com-
plement (anticodon) of the mRNA trinucleotide (codon) which codes for a specific amino acid. Genetic experiments have firmly established that the code word is a triplet. Each amino acid is designated by a particular trinucleotide sequence, and the codons do not overlap along the polynucleotide chain. Thus, the four-letter code (A, C, G, U) is translated into the approximately twenty-word “language” of amino acids out of which protein “sentences” are assembled. There are 64 possible combinations of the four-letter language and, since it is known that more than one codon can specify a particular amino acid, the code is “degenerate.”

mRNA functions in protein synthesis only when bound to ribosomes. Ribosomes are submicroscopic particles found in the microsomal fraction of cell homogenates obtained by differential centrifugation. In higher organisms, these particles are composed of approximately 50 per cent RNA and 50 per cent protein. The RNA:protein ratio in bacteria is approximately 60:40. As with messenger and soluble RNA’s, ribosomal RNA is assembled at specific regions of the genetic material. In the cells of higher organisms, ribosomes occur free or bound to the external surface of the lipoprotein membranes comprising the tubules, vesicles, and cisternae of the endoplasmic reticulum. Ribosomes possess a complex structure. Ribosomal RNA’s, ribosomal subunits, and ribosomes are classified by their sedimentation constants (S values) which are calculated from their rates of sedimentation in the ultracentrifuge. The ribosomal RNA’s of the bacterium Escherichia coli combine with proteins to form 30S and 50S subunits (Fig. 7) which, in turn, combine to form 70S ribosomes. The 70S ribosomes associate with mRNA to form polysomes which are active in protein synthesis. In the polysomes, the 70S particles appear to be held together by mRNA molecules bound to their 30S subunits.

The essence of the relation between DNA and protein structure is that one gene controls the synthesis of one polypeptide chain. Thus, a comparative study of the structure-function relationships in homologous proteins from a variety of simple and complex organisms at various stages of development should elucidate nu-

merous aspects of the evolution of gene structure and the mechanism of the regulation of gene function. Investigations of this type are being pursued in numerous laboratories and significant advances in our knowledge are being achieved. For example, elegant studies of the hemoglobin molecule have greatly increased our knowledge of the interrelations between the structure and function of proteins and their genetic control. The enzyme lactate dehydrogenase (LDH) is also being extensively investigated and is providing much information along similar lines.

LDH is an oxidoreductase that catalyzes the interconversion of lactate and pyruvate (Fig. 8). The reaction is mediated through the cofactor nicotinamide adenine dinucleotide (NAD). The major function of the enzyme may be to maintain the supply of

---

**Fig. 8. The glycolytic pathway.**
NAD⁺ necessary for glycolysis during periods of relative anaerobiosis by sequestering H⁺ in the form of lactate (which occupies a metabolic dead end).

The LDH of a wide variety of organisms ranging in complexity from bacteria to mammals is a tetramer composed of four equal-sized subunits. The molecular weight of these subunits is approximately 35,000. In mammals, these subunits exist in at least two compositionally and structurally distinct homopolymeric types. It is obvious that random assortment of two kinds of subunits into all possible combinations of four yields five molecular forms of the composition shown in Fig. 9.

It has become abundantly evident in recent years that numerous proteins, including many enzymes, exist in several physically distinct forms within the cells of a single organism. The multiple molecular forms of enzymes have been termed "isozymes." For numerous proteins, the physical basis of their multiple molecular forms resides in their polymeric structure.

LDH is ubiquitously distributed in nature in multiple molecular forms. Most mammalian tissues contain five principal LDH isozymes. The isozymes of this enzyme exhibit both species- and tissue-specific patterns (Figs. 10 and 11, respectively). Furthermore, during the course of embryonic development, the tissue-specific patterns undergo profound changes. This has been clearly demonstrated by the direct analysis of tissues of the mouse at different stages of development as illustrated in Fig. 12. LDH-5 is the predominant isozyme in all embryonic mouse tissue. As development progresses, the isozyme pattern shifts so that in most tissues, as exemplified by the heart, as increasing proportion of enzyme activity becomes localized in LDH-1. Only in those adult tissues, such as liver and skeletal muscle, in which LDH-5 is the predominant isozyme is the redistribution of enzyme activity during development relatively less dramatic.

From these studies there was no indication that the isozyme patterns of the different tissues shift synchronously. Also, it was obvious that the redistribution of enzyme activity did not occur to the same extent in all tissues. In some, such as mouse heart muscle, the change in isozyme pattern was too rapid to have resulted from a change in cell population. This strongly suggests that, during development, isozyme patterns change within individual cells.

From elementary genetic considerations, since the A and B subunits of LDH are different proteins, they must be under the control of different genes. Recent genetic evidence bears this out. An LDH mutant has been discovered in the deer mouse, Peromyscus maniculatus. In these animals, the mutation occurred at the B locus and, as theory predicts, the heterozygote produced 15 isozymes. Screening of diverse human populations has uncovered mutants at both the A and B loci but, to our knowledge, no double heterozygotes have yet been reported.

A third gene controlling the synthesis of a third type of LDH subunit—designated the C subunit—was discovered by Blanco and Zinkham. C polypeptides appear to be formed exclusively in the sperm. Isozymes containing C polypeptides are responsible for the so-called "X-bands" of LDH activity found on zymograms of testis homogenates. In some mammals only one
Fig. 10. Zymogram (starch gel electropherogram) illustrating the skeletal muscle LDH pattern of a representative of each vertebrate class. Enzyme preparation, electrophoresis, and staining performed according to Markert and Ursprung. Mammal, man (Homo sapiens). Bird, Adelie penguin (Pygoscelis adeliae). Reptile, rattlesnake (Crotalus adamanteus). Amphibian, Congo eel (Amphiuma means). Fish (Actinopterygii), fluke (Paralichys dentatus). Shark (Elasmobranchii), sand shark (Carcharias taurus). Lamprey, Cyclostomata (Pteromyzon marinus).

Fig. 11. Zymogram illustrating the tissue-specific patterns of LDH isozymes in bat (Myotis velifer) tissues.

X-band is observed, and it is assumed to be a tetramer of the C type subunits. Several X-bands have been detected in testis homogenates of other mammals. However, in these cases, it has been shown that the additional bands are the result of the polymerization of the C subunits with either A or B subunits. More recently Blanco, Zinkham, and Kupchyk have shown that the C gene of pigeons exists in two widely distributed allelic forms designated C and C' and that individuals fit into one of three phenotypic classes designated CC, C'C, and C'C'.

Although it is theoretically possible to form fifteen isozymes from three different subunits, no such number has been observed in sperm homogenates. The following interpretations of this may be brought forth. It is possible that the necessary freedom for random reassociation does not exist within the cell or that the gene controlling C polypeptide biosynthesis is turned on only when the A and B genes are turned off. It is also possible that certain hybrid molecules cannot be formed for purely physical reasons or that particular hybrid combinations are inactive. However, a mixture of A, B, and C subunits will readily recombine in vitro to yield the expected 15 different tetramers, eliminating the latter two possibilities from consideration.

The occurrence of isozymes of LDH in
nearly all vertebrates which have been examined strongly suggests that, for certain enzymes, multiplicity of form is evolutionarily advantageous and does not represent simple heterogeneity with no biological value. The implication that the individual isozymes subserve a specialized role in the economy of the organism is supported by the fact that, although all isozymes of LDH catalyze a characteristic chemical reaction, they possess markedly different physical and chemical properties. In the light of this evidence, it seems reasonable to conclude that isozymes are groups of molecules of common origin which have become differentiated to meet highly specific requirements within the cell.

The evolutionary history of LDH and perhaps that of all heteropolymeric proteins may have been similar. Originally, it is hypothesized, there was a single type of parental peptide whose synthesis was controlled by a single gene. This peptide associated into functional homopolymeric tetramers. In time, the gene became duplicated by unequal crossing over or some other mechanism. Simple duplication of itself would alter neither the structure of the gene nor that of its end product. Eventually, however, a mutation occurred in one of the two identical genes. This resulted in the production of two distinct polypeptides which on random tetramerization gave rise to heteropolymers and,
thus, to multiple molecular forms of the enzyme. The initial mutationally induced structural alternations of the gene did not alter the structure or functioning of its polypeptide end product significantly. However, with the selective accumulation of further mutationally induced alterations in both genes, their polypeptide end products became highly specialized for functioning under different environmental conditions or in different locations within the cells comprising the tissues of the organism. Evidence in support of this has been brought forth by several investigators.

The fundamental problem of contemporary developmental genetics, the regulation of gene action, is enormously complex and cannot be solved by one approach. Studies of differential isozyme synthesis during development can be utilized as highly sensitive monitors of changing gene function and, employed in conjunction with other kinds of investigations, may lead to increased insight into the molecular mechanisms regulating gene function.

REFERENCES


