CHAPTER

8

Regulation of Gene Action

Many invertebrate animals – and some vertebrates – change form as they mature. The free swimming larva (top) grows into a slow – moving bottom – dwelling starfish (above). Changes in gene activities cause this metamorphosis. The pattern of change is itself controlled by other genes.

The synthesis of particular gene products is controlled by mechanisms collectively called **gene regulation**. Evidently, though cells contain the genetic capacity for the synthesis of an enormous number of different products (proteins), not all of these products are present at any given time, many being selectively activated only upon special occasion and in response to some environmental stimulus. For example, in prokaryotes some enzymes are synthesized **constitutively** (*i.e.*, continuously), indicating that transcription of mRNA is constantly occurring in them. However, other enzymes are synthesized only when a need for their action arises, and when this need has been fulfilled, enzyme synthesis stops. Transcription of mRNA in this case is evidently initiated only on demand and must, therefore, be subject to regulation. Exhaustive investigations have established that regulation of gene activity both in prokaryotes and eukaryotes may occur at three levels: transcription, translation and post-translation (*i.e.*, folding and processing of proteins).

The gene regulatory systems of prokaryotes and eukaryotes are slightly different from each other. Prokaryotes are generally free-living unicellular organisms that grow and divide continually as long as environmental conditions are suitable and the supply of nutrients is adequate. Thus, their gene regulatory systems are adapted to provide the maximum growth rate in a particular environment, except when such growth would be deterimental. This procedure seems to apply

DNA swquences involved in gene regulation have been identified

to the free-living unicellular eukaryotes such as yeast, algae and protozoa.

The demands of tissue-forming eukaryotes are quite different from those of prokaryotes. For example, in a developing organism, in an embryo, a cell must not only grow and produce many progeny cells but also must undergo considerable change in morphology and biochemistry (*i.e.*, become differentiated) and then maintain the changed state. Also, during the growth and cell-division phases of the organism, these cells are challenged less by the environment (in contrast to the bacteria), because the composition and concentration of their growth media do not change drastically with time. Some examples of such media are blood, lymph, or other body fluids, or in the case of marine animals, the sea water. Lastly, in an adult organism, growth and cell division in most cell types have stopped, and each cell needs only to maintain itself and its properties. Many other examples could be quoted, the main point is that because a typical eukaryotic cell faces different emergencies than a bacterium does, the gene regulatory mechanisms of eukaryotes and prokaryotes are not the same.

REGULATION OF GENE ACTION IN PROKARYOTES

In prokaryotes (*e.g., E.coli*), the activities of genes are regulated according to the following mechanisms :

1. Transcriptional Control Mechanisms

In bacteria, there occur several mechanisms of gene regulation at the level of transcription. A notable method depends on whether the enzymes being regulated act in catabolic (degradative) or anabolic (synthetic) metabolic pathways. For example, in a multistep catabolic system the availability of the molecule to be degraded commonly determines whether the enzymes in the pathway will be synthesized. In contrast, in a biosynthetic pathway the final product is often the regulatory molecule. Even when a single protein molecule is translated from a monocistronic mRNA molecule, the protein may be **autoregulated**, *i.e.*, the protein itself may inhibit initiation of transcription and high concentrations of the protein may cause less transcription of the mRNA that encodes the protein. The molecular mechanisms for each of the regulatory patterns differ greatly and are of the following two types— negative regulation and positive regulation (Fig. 8.1). In a **negative regulated system**, an inhibitor is present in the cell and prevents transcription. An antagonist of the inhibitor, called an **inducer**, is needed to allow initiation of transcription. In a **positively regulated system**, an **effector** molecule (which may be protein, a small molecule, or a molecular complex) activates a promoter; no inhibitor must be countermanded. Negative and positive regulation are not mutually exclusive, and some systems are both positively and negatively regulated, utilizing two regulators to respond to different conditions in the cell. Thus, a catabolic system may be regulated positively or negatively. In a biosynthetic (anabolic) pathway, the final product usually regulates negatively its own synthesis; in the simplest type of negative regulation, absence of the product increases its synthesis and presence of the product decreases its synthesis.

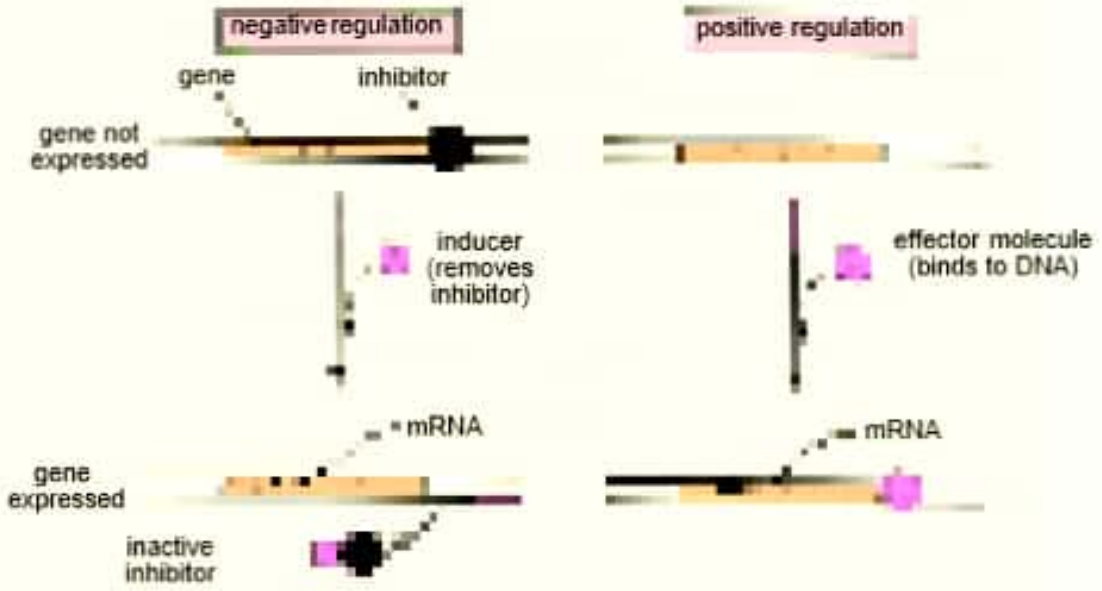


Fig. 8.1. The distinction between negative and positive regulation. In negative regulation an inhibitor, bound to the DNA, must be removed before transcription can occur. In positive regulation an effector molecule must bind to the DNA (after Freifelder, 1985).

A. Negative Control

(i) Inducible Operons (Inducible Systems)

An inducible enzyme is produced only when its substrate (inducer) is present in the environment (*i.e.*, active repressor + inducer = inactive repressor). Most enzymes in this category are catabolic in their activity. **F. Jacob** and **J. Monod** in 1961, on the basis of their study on the inducible system for the synthesis of β- galactosidase enzyme in *E.coli*, proposed a model in order to explain the induction or repression of enzyme synthesis. This model is popularly known as **operon model** and has been variously modified, ever since it was originally proposed by **Jacob** and **Monod**. According to this model, an **operon** was defined as a unit of coordinated control of protein synthesis which consisted of (i) an **operator gene** which controlled the activity of (ii) a number of **structural genes** which took part in the synthesis of protein(s). This means that the structural genes will synthesize mRNA under the operational control of an operator gene. The operator gene, in its turn, is under the control of a repressor molecule synthesized by a **regulator gene**, which is not a part of the operon. Thus, the members of an operon are transcribed coordinately a single, long, polycistronic mRNA molecule. One such operon in *E.coli*, called the **lactose** or **lac operon**, has provided a model system for the study of gene regulation.

Fancois Jacob and Jacques Monod.

Mechanism of lac operon. A *lac* operon contains three structural genes or cistrons, namely *z*, *y* and *a*, whose products (=enzymes) are involved in the breakdown (catabolism) of the sugar lactose (Fig. 8.2). Gene *z* contains 3063 base pairs and codes for an enzyme, **β- galactosidase**, which converts lactose into glucose and galactose; while gene *y*

contains 500 base pairs and determines the structure of an enzyme, **galactoside permease** which is a plasma-membrane bound protein and facilitates the entrance of lactose into the cell. Gene *a* comprises 800 base pairs and specifies an enzyme, **thiogalactoside acetylase**, which transfers an acetyl group from acetyl-CoA to β- galactoside (*i.e.*, this enzyme is indirectly involved in lactose utilization). Mapping experiments employing mutations of these three genes have demonstrated that the gene order is *z-y-a*.

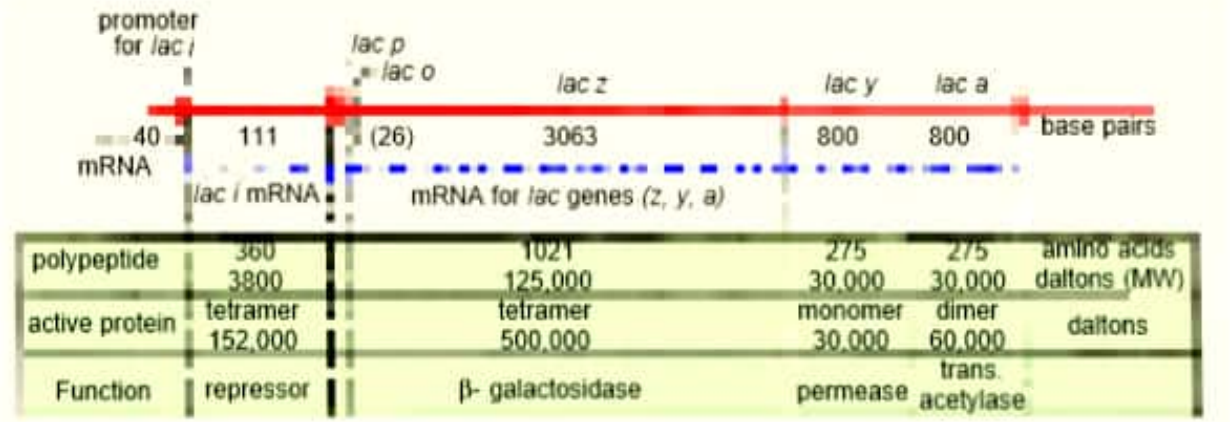
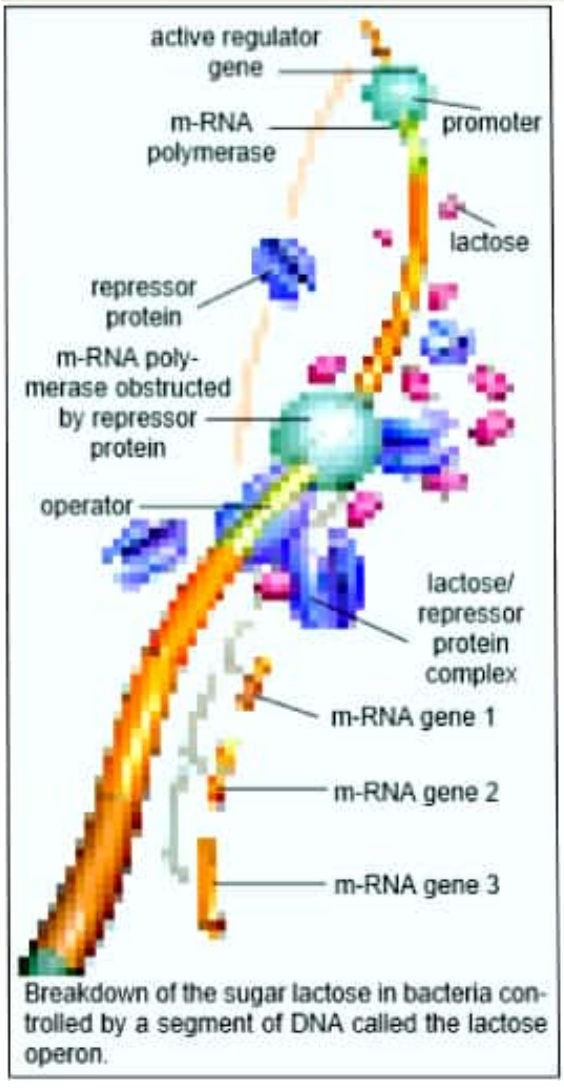


Fig. 8.2. The lactose operon of *E.coli* and its regulatory gene (after Lewin, 1990).



Breakdown of the sugar lactose in bacteria controlled by a segment of DNA called the lactose operon.

Normally, the synthesis of these three enzymes is not constitutive and in the absence of lactose only a few molecules of each enzyme are present. However, when lactose is provided, all three enzymes are synthesized rapidly and simultaneously as co-ordinated response to the presence of this substrate. Thus, lactose acts to **induce** the production of the enzymes needed for its catabolism. Since all three enzymes are synthesized through the translation of a single polycistronic mRNA, it follows that the entire operon is responding as a unit to the presence of inducer.

A clue to the cause of this controlled response on the part of the *lac* operon was provided by a mutation in whose presence all three enzymes were produced constitutively regardless of the presence or absence of lactose. Since the enzymes themselves were normal, it was concluded that this mutation had occurred in a so far unidentified controlling element which was given the name **operator** (**o**). By mapping experiments the operator element was found to be situated immediately adjacent to gene *z*, so that the overall gene order was *o-z-y-a*. The operator is found to comprise 26 base pairs.

Another evidence to the mechanism controlling the *lac* operon was provided by the discovery of a gene called **inducer (i)** or **regulatory gene** whose locus was closely linked to the operon, but was separated from that of the operator, gene order being *i-o-z-y-a*.

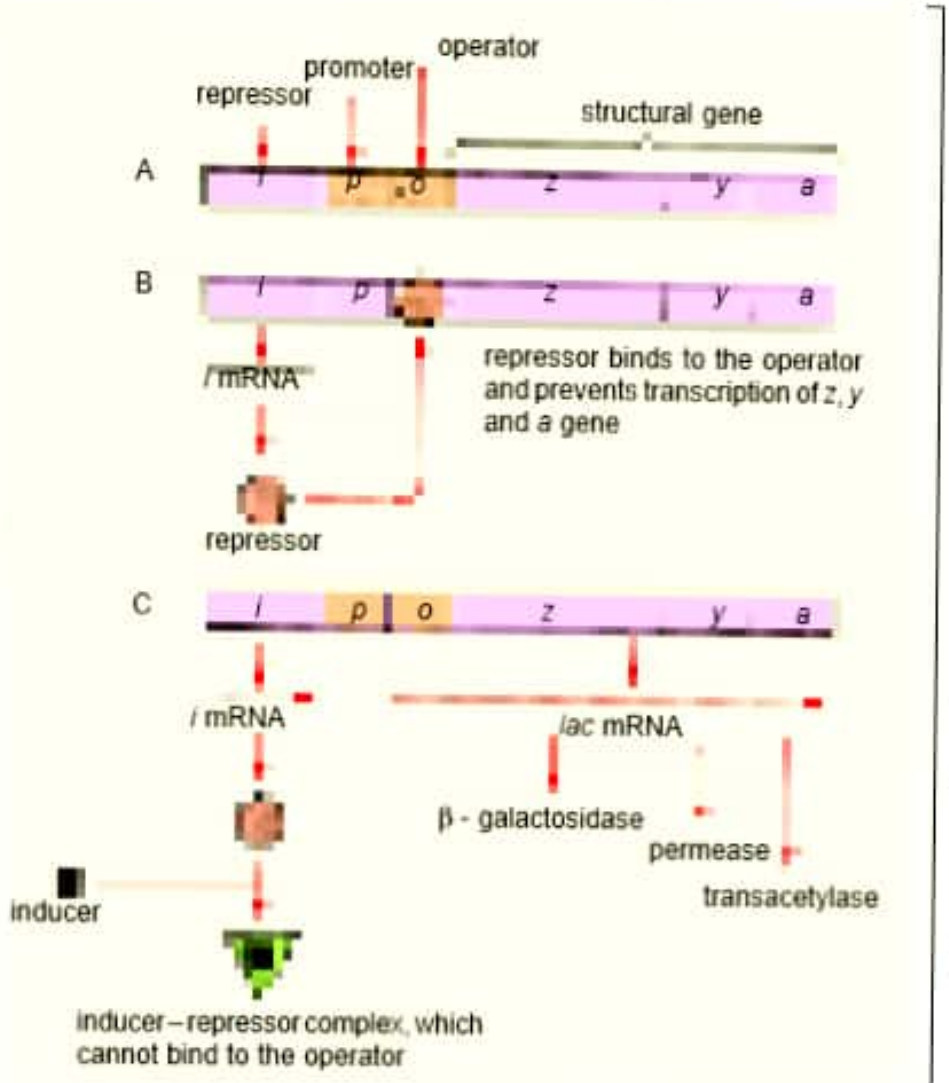


Fig. 8.3. A–Genetic map of the *lac* operon, not drawn to scale: the *p* and *o* sites are actually much smaller than the genes; B–*Lac* operon in repressed state; C–*Lac* operon in induced state. The inducer (lactose) changes the shape of repressor, so repressor can no longer bind to the operator (after Freifelder, 1985).

The *i* gene specified a product (called **repressor**) that could diffuse from the site of transcription and translation to the altogether different region of the genome and there influences the function of the wild type operator (*o*+) and *z*+ gene. Gene *i* is 1111 base pairs long and is transcribed separately from the genes of the *lac* operon as a monocistronic mRNA which is translated on the ribosomes into repressor protein. The repressor protein is a diffusible tetramer protein having 152,000 dalton M.W.

To clarify the respective roles of these elements, **F. Jacob** and **J. Monod** in 1961 proposed that a distinction be made between structural genes such as *z*, *y* and *a* that designate proteins required for cell metabolism, and regulatory genes such as *i*, whose products participate in control mechanisms imposed on structural genes. The operator element was designated a **controlling site** which governed the transcription of physically adjacent structural genes. **Jacob** and **Monod** also proposed a theory to explain the interaction of these loci in the regulation of coordinate enzyme synthesis. According to this theory, the regulatory gene *i* specified a repressor protein which in the absence of the inducer (lactose), was bound to the operator (*o*), thereby inactivating the operator and preventing transcription of the three *lac* cistrons (Fig. 8.3). Induction of transcription was explained as the result of the binding of the inducer (= lactose) to the repressor protein such that the repressor dissociated from the operator. Upon this release from repression, the operator would permit the transcription of the adjacent operon and coordinate enzyme synthesis would follow. The repressor protein was visualized as a molecule with two different, non-overlapping binding sites, one for the operator and the other for the inducer. Union of repressor with inducer would cause a change in the conformation of the repressor protein which rendered the binding site for the operator non-functional. However, when the supply of inducer was depleted through the activity of *lac* enzymes, dissociation of the inducer-repressor complex would occur, permitting a reverse change in conformation so that the repressor could once more bind to the operator to shut down transcription. Such a kind of reversible change in the conformation of a molecule is called **allostery**.

When **Jacob** and **Monod** conducted their studies in 1960s, the existence of the promoter and the mechanism of repressor function were not known. As already described in Chapter 5, the promoter (*p*)is the site of RNA polymerase attachment. Mutation of the promoter causes a decrease in enzyme synthesis. The complex formed between the DNA of the promoter and RNA polymerase serves to initiate strand separation and sense strand selection, so that mRNA synthesis can begin when the start codon within the operator is reached. The *p* site has been mapped and has been found to lie between *i* and *o* genes. We should note that gene *i* has its own promoter (*p'*) located prior to the *i* locus.

(ii) Repressible System

Some enzymes are normally present in the cell but cease to be synthesized when high concentrations of their end product are present. Such enzymes are called **repressible enzymes**; while the end product is called **corepressor**. A gene called **regulator gene** produces a substance called **aporepressor** which unites with corepressor to form a functional **repressor** molecule. This repressor molecule or substance inhibits synthesis of mRNA by all genes specifying enzymes in the synthetic pathway. Most repressible enzymes are found in anabolic pathways. Hormones may exert their phenotypic effects by derepressing gen-es previously repressed.

Example. In *Salmonella ty-phimurium* when the amino acid, histidine occurs in high concentration in the medium, that starts to act as corepressor. As a corepressor, this amino acid terminates the synthesis of 10 enzymes which are required in pathway to histidine. This kind of repression is called **coordinate repression** (Fig. 8.4).

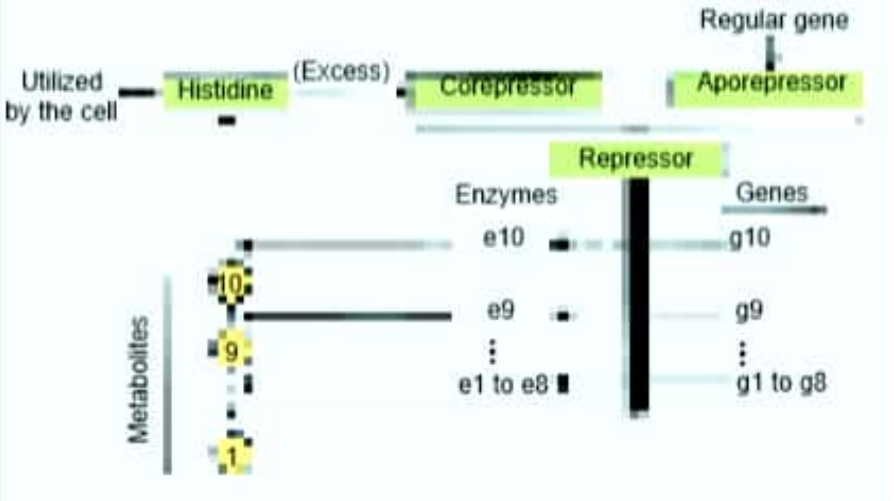


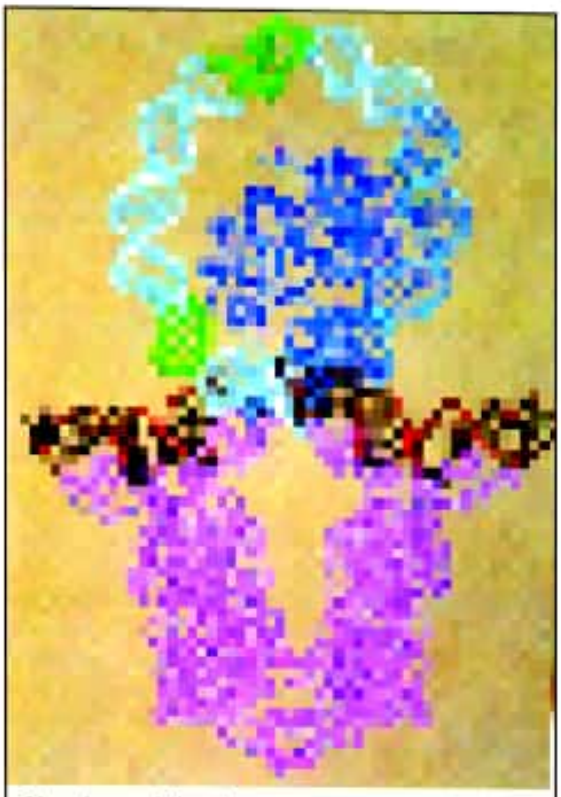
Fig. 8.4. Coordinate repression in *Salmonella typhimurium*.

B. Positive Control

Many bacterial genes are under positive control. This mode of gene regulation is attributed to the presence of factors that enhance the attachment of RNA polymerase to the promoters and initiation of mRNA synthesis.

(i) Effects of glucose on lac operon (Catabolic repression) or Glucose effect. The function of β-galactosidase enzyme in lactose metabolism is to form glucose by cleaving lactose. (The other cleavage product, galactose, is also ultimately converted to glucose by the enzymes of galactose operon). Thus, if both glucose and lactose are present in the growth medium, activity of *lac* operon is not needed, and indeed, no β- galactosidase is formed until virtually all of the glucose in the culture medium is consumed. The lack of synthesis of β-galactosidase is a result of lack of synthesis of *lac* mRNA. No *lac* mRNA is made in the presence of glucose, because in an addition of an inducer to inactivate the *lac i* repressor, another element (*i.e.*, cAMP-CAP) is needed for initiating *lac* mRNA synthesis; the activity of this element is regulated by the concentration of glucose. However, the inhibitory effect of glucose on expression of the *lac* operon is quite indirect.

Small molecules, the **cyclic AMP** (**cAMP**), are present in *E.coli* and many other bacteria. Cyclic AMP is synthesized enzymatically by **adenyl cyclase** and its concentration is regulated indirectly by



Structure of the *lac* operon repression loop. The *lac* repressor, shown in violet, binds to two DNA regions (red) and immediately upstream from the CRP binding site, within the loop is the CRP binding site (medium blue), shown bound with CAP protein (dark blue).

system. The cAMP-CAP complex must be bound to a base sequence in the DNA in the promoter region in order for transcription to occur (Fig. 8.5). Thus, cAMP-CAP is a positive regulator, in contrast with the repressor, and *lac* operon is independently regulated both positively and negatively.

(ii) Tryptophan operon. The tryptophan (*trp*) operon of *E.coli* is responsible for the synthesis of the amino acid tryptophan. Regulation of this operon occurs in such a way that when tryptophan is present in the growth medium, *Trp* operon is not active. That is, when adequate tryptophan is present, transcription of the operon is inhibited; however, when its supply is insufficient, transcription occurs. The *Trp* operon is quite different from the *lac* operon in that tryptophan acts directly in the repression system rather than as an inducer. Moreover, since the *Trp* operon encodes a set of biosynthetic (or anabolic) rather than catabolic enzymes, it is subject to repression control.

glucose metabolism. When bacteria are growing in culture medium containing glucose, the cAMP concentration in the cells is quite low. In a medium containing glycerol or any carbon source that cannot enter the biochemical pathway used to metabolize glucose (the glycolytic pathway), the cAMP concentration becomes high. The mechanism by which glucose controls the cAMP concentration is poorly understood; the significant point is that cAMP regulates the activity of the *lac* operon (and other several operons as well). *E. coli* contains a protein called the **catabolic activator protein (CAP)**, which is encoded in a gene called *crp*. Mutants of either *crp* or the adenyl cyclase gene are unable to synthesize *lac* mRNA, indicating that both CAP and cAMP are required for *lac* mRNA synthesis. CAP and cAMP bind to one another to form a unit, called **cAMP-CAP** which is an active regulatory element of the *lac*

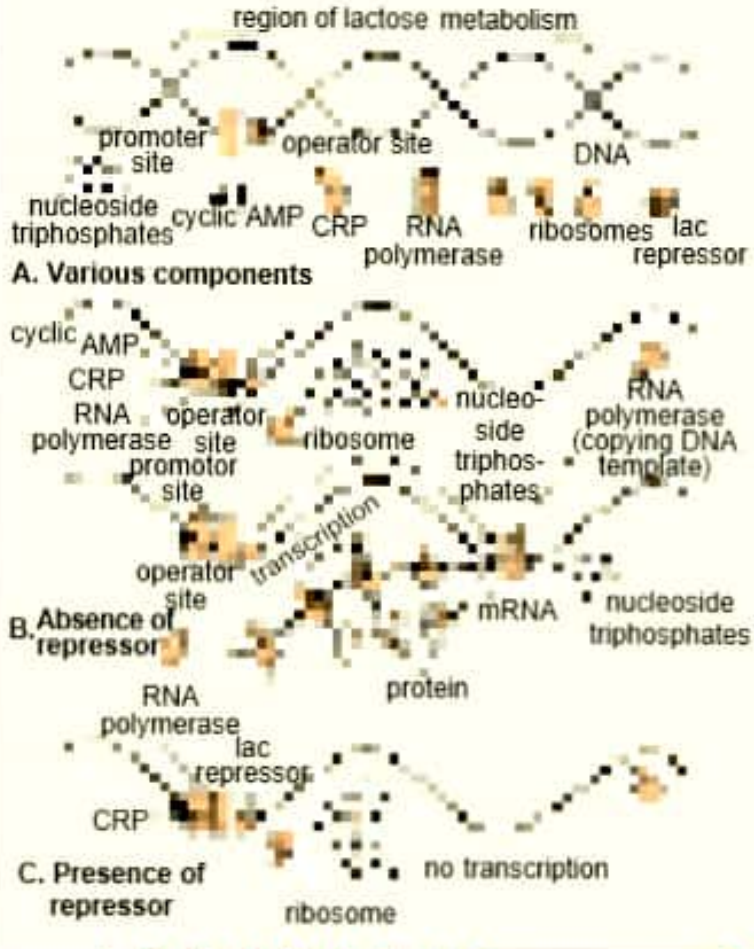


Fig. 8.5. Mechanism involved in the positive control system for the regulation of gene activity in *E.coli lac* operon. It should be noted that only in absence of the repressor, RNA polymerase enzyme can travel and transcribe *lac* operon as shown in B. The repressor when present on the operator site (C), it is an obstacle in the path of RNA polymerase.

bolic enzymes, neither glucose nor cAMP-CAP has a role in the operon activity.

A simple on-off system, as in the *lac* operon, is not ideal for a biosynthetic pathway; a situation may arise in nature in which some tryptophan is available, but not enough to allow normal growth if synthesis of tryptophan was totally shut down. Tryptophan starvation is prevented by a *modulating system in which the amount of transcription in the derepressed state is determined by the concentration of tryptophan*. Such a mechanism is found in many operons responsible for amino acid biosynthesis.

Tryptophan is synthesized in five steps, each requiring a particular enzyme. In the *E.coli* chromosome the genes encoding these enzymes are located adjacent to one another in the same order as they are used in the biosynthetic pathway; they are translated from a single polycistronic mRNA molecule. These genes are called *TrpE*, *TrpD*, *TrpC*, *TrpB* and *TrpA*. The *TrpE* gene is the first one translated. Adjacent to the *TrpE* gene are the promoter, the operator and two regions called the **leader** and the **attenuator**, which are designated *TrpL* and *Trpa* (not TrpA) respectively (Fig. 8.6). The repressor gene *TrpR* is located quite far from gene cluster.

The regulatory protein of the repression system of the *Trp* operon is the *Trp R*-gene product. Mutations either in this gene or in the operator cause constitutive initiation of transcription of trp-mRNA as the *lac* operon. This regulatory protein is called *Trp* **aporepressor** and it does not bind to the operator unless tryptophan is present. The aporepressor and the tryptophan molecule join together to form the active *Trp* repressor which binds to the operator. The reaction scheme is as follows :

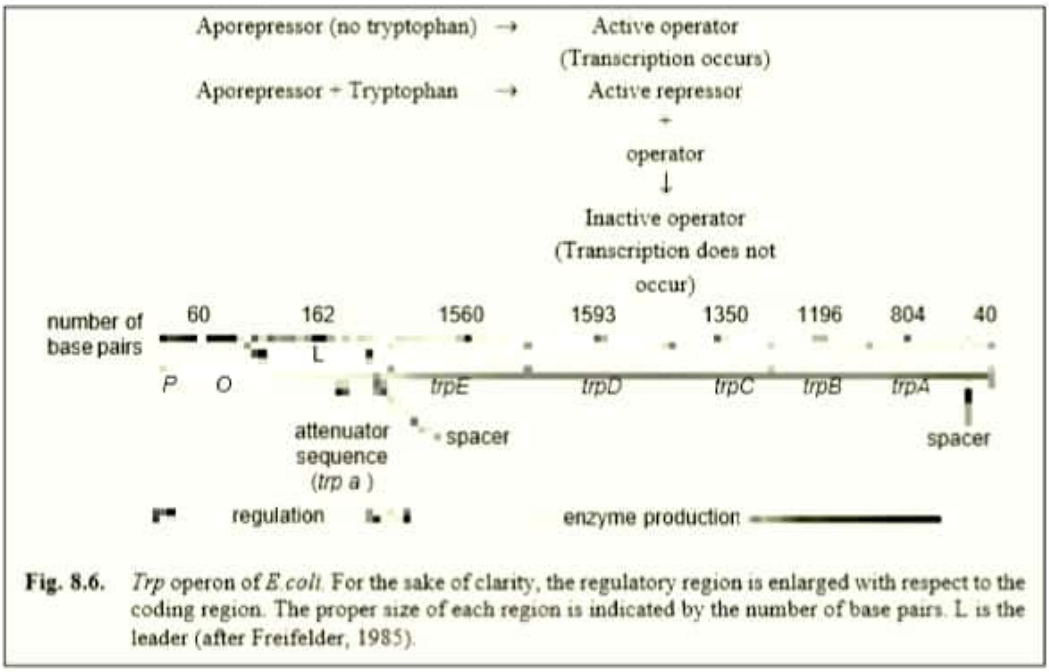


Fig. 8.6. *Trp* operon of *E.coli*. For the sake of clarity, the regulatory region is enlarged with respect to the coding region. The proper size of each region is indicated by the number of base pairs. L is the leader (after Freifelder, 1985).

2. Translational Control

In prokaryotic gene regulation at the translation level, the lifetime of a mRNA molecule may be genetically determined. Enzymatic degradation of mRNA is from the 5' to the 3' end, *i.e.*, the end of the RNA that is first synthesized is also the end that is first degraded. The average lifetime of many mRNA molecules of *E.coli* is only two minutes at 37°C. The specific nucleotide sequences at the 5' end may influence its susceptibility to enzymatic digestion. Further, catabolic enzymes are denied access to the mRNA when the ribosome coated at their 5' ends (*i.e.*, in case of polyribosomes). Hence, the lifetime of mRNAs may also be correlated with the number of free ribosomes available at any given

moment to translate mRNA molecules. Bacteria vary their rates of protein synthesis by varying their ribosomal content rather than by varying the translational rate.

Example. In the lactose system of *E.coli*, there are three structural genes under control of a common operator locus determining production of (1) β- galactosidase, (2) galactoside permease and (3) galactoside acetylase. These three proteins are produced in the respective ratios 1:1/2:1/5, reflecting their respective locations relative to the 5' (operator) end of the polycistronic mRNA in which they are coded (these differences are the examples of translation regulations). Thus, there is a *polarity gradient* within the polycistronic mRNA that reduces the probability of cistron translation as a function of its distance fro the 5' end. It is hypothesized that ribosomes attach to different starting points (ribosome-binding sites) along the polycistronic mRNA at different rates as reflected by the relative amounts of the three proteins synthesized.

3. Post-translation Control (Feedback Inhibition or End Product Inhibition)

The expression of genes also can be regulated after proteins have been synthesized. This is called **post-translational control of gene action**. Feedback inhibition is a regulatory mechanism which does not affect enzyme synthesis, but rather inhibits enzyme activity (Fig. 8.7). The end product of a biosynthetic pathway may combine loosely (if in high concentration) with the first enzyme in the pathway. This union does not occur at the catalytic site, but it does modify the tertiary structure of an enzyme and, hence, inactivates the catalytic site. This **allosteric transition** of protein molecule blocks its enzymatic activity and prevents overproduction of end products and their intermediate metabolites.

Example. The studies on isoleucine synthesis in *E.coli* (**Umbarger** 1961) demonstrated that addition of isoleucine (the end product of a five step conversion of threonine) to a culture of the bacteria resulted in immediate blocking of the threonine→isoleucine pathway. In the presence of added isoleucine, the cells preferentially use this **exogenous** end product (*i.e.*, isoleucine) and their own isoleucine synthesis becomes ceased. Moreover, the production of each of the five enzymes is not interfered with, but action of an enzyme responsible for deamination of theronine to α-ketobutyrate is inhibited by the end product, isoleucine.

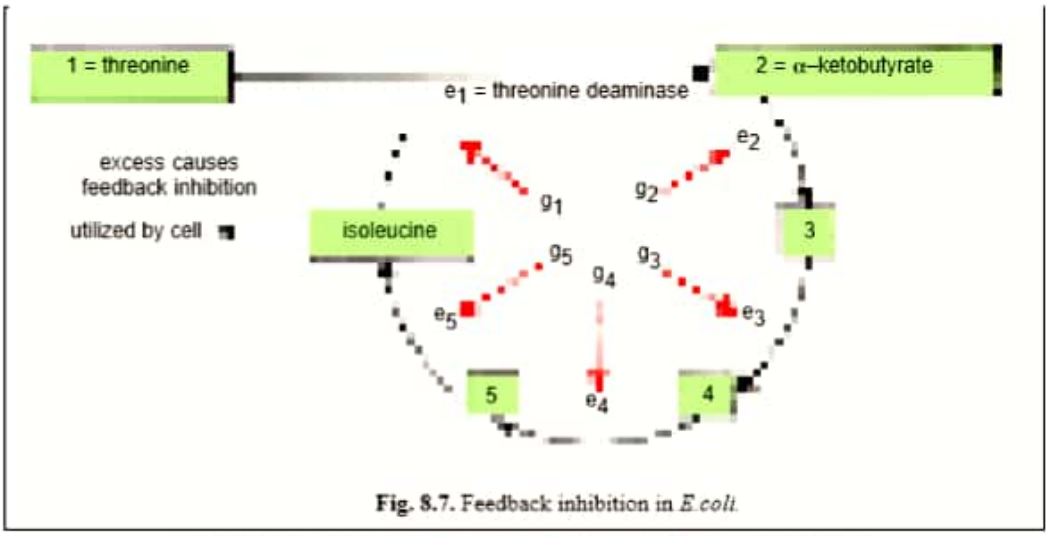


Fig. 8.7. Feedback inhibition in *E.coli*.

REGULATION OF GENE ACTION IN EUKARYOTES

In eukaryotes the following two kinds of controls or regulations of gene expression occur : 1. **Short-term** or **reversible regulation** corresponds to the kind of regulation we studied in bacteria and it represents a cell's response to fluctuations in the environment, specifically, it involves changes in

activities or concentrations of enzymes as particular substrates and or hormone levels rise and fall. The changes a cell experiences during a cell cycle, particularly the fluctuations in rates of DNA, RNA, and protein synthesis that regularly occur with respect to the time of mitosis can also be placed in this category. 2. **Long-term** or **irreversible regulation** includes the phenomena associated with **determination**, **differentiation**, or more generally development: it is involved in the numerous steps by which a fertilized egg becomes an organism of, perhaps, trillions of cells with diverse and ultimately quite permanent roles to play in the maintenance of the whole. Short-term regulations also occur in developing and differentiating eukaryotic cells side by side of long-term regulation.

Both of these types of regulations of gene activities in eukaryotes, now, are considered to occur at the following levels involving diverse mechanisms: 1. Regulation at the level of DNA; 2. Regulation at the level of transcription; 3. Regulation at the level of translation; and 4. Regulation at the level of post-translation.

A. Regulation of Gene Action at the Level of Genome

In eukaryotic cells, it seems that certain classes of genes are transcribed more or less continuously, and only in extreme situations their activities are repressed. For example, genes coding for larger ribosomal RNA (28 S or 18 S rRNA) or transfer RNA (tRNA) are present as multiple copies forming **simple multigene family**. These genes are transcribed uniquely by RNA polymerase I for the larger ribosomal RNA or by RNA polymerase III for tRAN and 5S RNA. Although the products of some of these genes, the ribosomes, are used continuously in all cells, it does not conform that all of these multiple copies are continuously transcribed at maximum rate. Electron micrographs of spread chromatin from nucleoli often show that some of the repetitious rRNA genes are inactive. It is also true that in the nucleated erythrocytes of lower vertebrates such as *Xenopus*, all genes may be turned off (**Maclean** *et al.*, 1972), including those for ribosomal RNA and tRNA. Therefore, it is clear that mechanisms do exist for inactivating sequences even those regarded to be constitutive in normal cells.

A giant chromosome of a midge showing puff.

Some of the clearest demonstrations that some specific genes are at least available for transcription in different kinds of differentiated cells are provided by *Drosophila* and other organisms (Fig. 8.8). For example, the pattern of bands and interbands of **polytene chromosomes** of *Drosophila* does not vary between different larval tissues, yet it is now concluded that the interband regions probably represent **housekeeping genes** which code for essential proteins, that they are expressed in every cell, and that they are retained in a state of permanent decondensation (**Bautz** and **Kabisch**, 1983). Thus, 'housekeeping' genes may be 'left on' for much of the life of the cell when transcription of even the most essential housekeeping genes ceases (*i.e.*, during mitosis).

Further, when we consider the case of the "**cell-specific**" **genes**, **luxury genes** or **smart genes**, which code for the products only found in specialized tissues, it becomes immediately clear that differential expression is the rule. Whether expression of gene is measured at the level of the messenger RNA or the protein, genes coding for products such as globin, crystallin, fibroin, ovalbumin, casein and immunoglobulin give every indication of complete repression in all but the specialized tissue characterized by their presence.

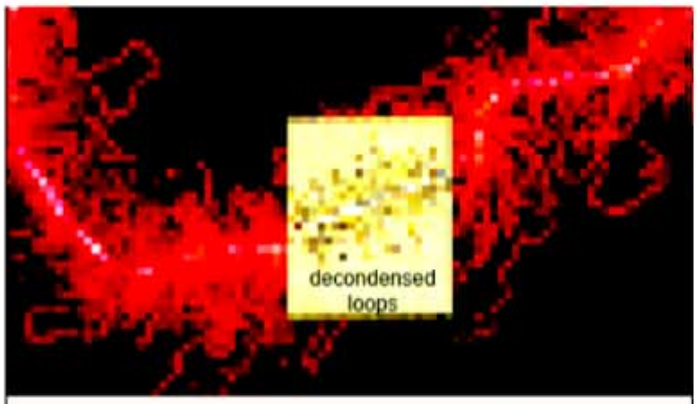
Thus, at the level of genome (*i.e.*, DNA), the following five modes of regulation are operative:

(i) Situations of total genetic shutdown. (a) During mitotic phase of the cell cycle, chromatin is highly condensed to form chromosomes, and transcriptional activity of all genes is suspended.

(b) During meiotic division of germ cells a somewhat similar situation to (a) is evident, although in some rare cases, such as **lampbrush chromosomes** of meiotic diplotene in vertebrates (**Vlad**, 1983), transcription proceeds very actively.

(c) The nucleus of mature nucleated erythrocytes of amphibians is transcriptionally inactive. Chromatin in these cells is highly condensed but not organized into discrete chromosomes (**Chegini** *et al.*, 1981). However, transcription can be partially reactivated in these nuclei by transferring them into new cytoplasm or exposing them *in vitro* to altered environmental conditions.

(d) In mammalian females, one of the two X chromosomes present in somatic cells undergoes condensation in early embryonic stages to become heterochromatic **sex chromatin** or **Barr body** (Dosage compensation). A variety of experiments indicate that most, though not all, genes of the condensed X chromosome are turned off. In developing oocytes, as opposed to somatic cells, Barr bodies are not present, the activities of both X chromosomes being required for normal oogenesis.



Lampbrush chromosome of a newt.

In somatic cells of normal XY males, genes of the single X chromosome remain active and Barr bodies are not found. However, in germ cells, the X chromosome is inactivated prior to spermatogenesis, otherwise, it may prevent sperm maturation and lead to sterility. In one extreme case, that of the creeping vole, *Microtus oregoni*, the X chromosome is eliminated from the germ cells of males by a special process of nondisjunction (see **Farnsworth**, 1988).

(e) Sperm cells clearly contain a complete genetic endowment but no transcription occurs until the sperm nucleus is activated within the egg cytoplasm.

(f) Complete suspension of transcriptional activity is also known in the following cases: cells of some plant seeds; cells within diapausing *Artemia* gastrulae; cells within inactive organisms such as desiccated *Tardigrada*; nuclei within bacterial and fungal spores; and nuclei within desiccated amoeba cells, as for example in the slime mould *Dictyostelium*.

(ii) Evidence for constitutive expression of some genes. (a) If the interbands of *Drosophila* polytene chromosomes are correctly interpreted as being loci for "housekeeping" genes, then the evidence is that such chromatin is permanently decondensed and is transcribed at a low but constant rate (**Semeshin** *et al.*, 1979).

(b) Electron microscopy of spread films of DNA extracted from nucleoli of *X. laevis* oocytes reveals tandemly arranged sequences coding for the 45S precursor of ribosomal RNA, each gene adorned with a Christmas-tree arrangement of RNA in the process of synthesis (*i.e.*, transcription).

(c) There is a constant and universal requirement for the products of certain genes in all cells and at all times. These include products such as the four kinds of rRNA—28S, 18S, 5.8S and 5S; tRNA of 20 basic types, and a few hundred proteins such as histones, ubiquitin and lactate dehydrogenase, RNA polymerase, and the like.

(iii) Many genes are expressed only in certain tissues. (a) *Xenopus* provides a good example of regulation of 5S genes. *Xenopus borealis* possesses 19,000 copies of the oocyte-specific 5S rRNA genes, and these genes are active only in the oocyte and in no other cell.

(b) The enzyme lactate dehydrogenase (LDH) is coded by a small family of genes, each gene determining the structure of a subunit. Subunits A and B are expressed in almost all mammalian cells, but one of the genes in the family, coding for subunit C, is active only in spermatocytes within the developing testes.

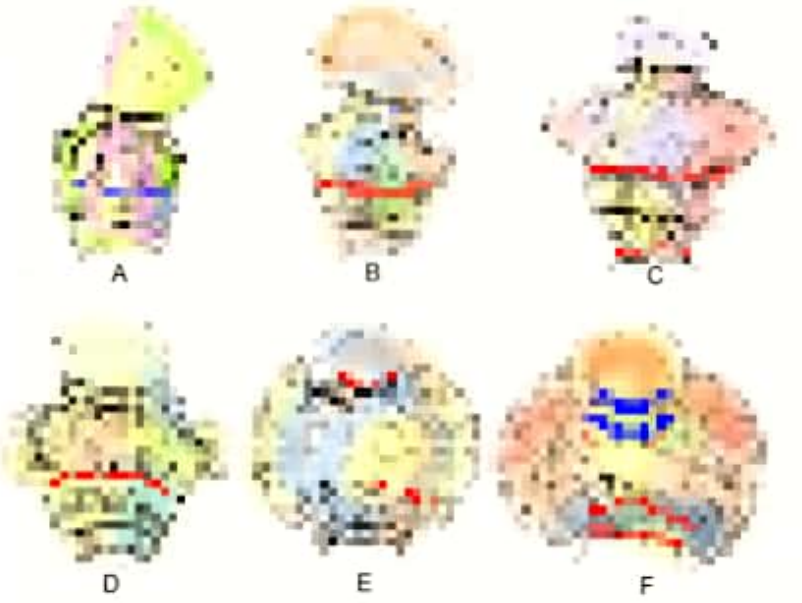


Fig. 8.8. The development of a chromosomal puff in a larval salivary gland cell nucleus of *Chironomus tentans*.

(c) The **puffing** (*i.e.*, chromatin decondensation) of restricted segments of the polytene chromosomes of *Drosophila, Chironomus* (Fig. 8.8), etc., provides visible evidence of the activity of genes coding for cell-specific products. Certain puffs, known as **heat-shock puffs**, can be induced to appear specifically when salivary glands are exposed to heat shock either *in vivo* or *in vitro*. The correlation between such chromatin decondensation and transcription activity is readily proved by autoradiography using tritiated precursors of RNA.

(iv) Some DNA is never transcribed in any cell. Analysis of various types of DNA sequences existing in eukaryotic cells reveals that some DNA is comprised of tandemly repeated short sequences that are concentrated in **heterochromatin** such as centromeres of chromosomes and the Y chromosome. Current evidence indicates that much of this DNA is never transcribed in any cell. Some spacer sequences occur between genes, for example, between multiple copies of genes for ribosomal RNA. Such spacer sequences are often taken to be untranscribed, but now it is found that some of the spacer DNA may transcribe nuclear RNA molecules.

The very large size of the genomes of the higher eukaryotes certainly indicates that much of the DNA is redundant (= repetitious) and probably not utilized as coding or regulatory sequence. The **pseudogenes** found in many gene families, often presumed to have arisen as cDNA (=complementary DNA) copies of reverse transcribed message. They lack introns and contain many stop signals so that RNA polymerase molecules fail to move very far along them.

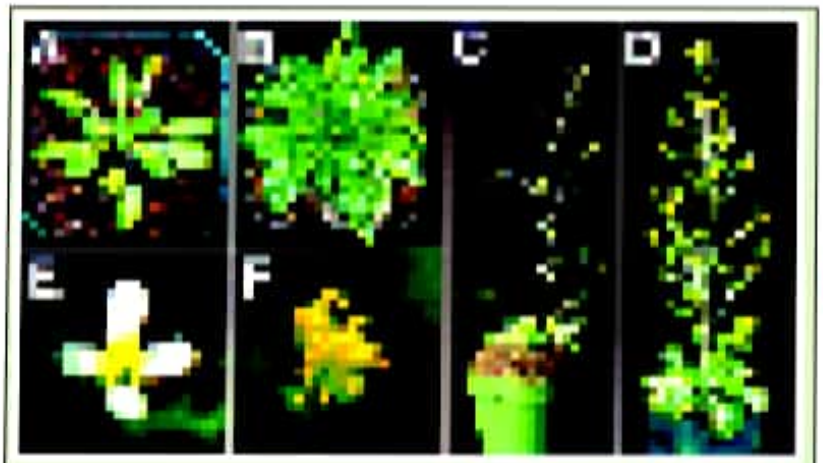
(v) Some DNA is spliced to cause gene rearrangement. Such a mechanism occurs during expression of immunoglobulin (Ig) genes.

B. Regulation of Gene Action at the Level of Transcription

(a) Chromatin reconstitution experiments. Chromatin has three main components—DNA, histones and non-histones. While it is known since early 1960s that histones may be involved in repressing gene activity, the specific regulation by non-histones was shown only during 1970s. **Gilmour** and **Paul** (1973) performed a chromatin reconstitution experiment to demonstrate the positive role of non-histones in regulation of gene activity (Fig. 8.9). They isolated the chromatin from

different tissues separately and then dissociated into DNA, histones and non-histones. This is followed by **chromatin reconstitution** using either the three components derived from the chromatin of the same tissue or by combining the non-histones of one tissue with the DNA and histones of another tissue (Fig. 8.9). From such experiments, it was demonstrated that the mRNA which is synthesized *in vitro* from reconstituted chromatin, mainly depended on the source of non-histone proteins (see **O'Malley** *et al.*, 1977).

(b) Change in chromatin conformation. Evidently nucleosomes continue to be present on most transcriptionally active DNA sequences, but they are probably reduced in number. Thus, although evidence from some laboratories suggests that the ribosomal genes of the amphibian nucleolus lack nucleosomes, active gene loci in *Drosophila* give a positive reaction to antibodies against H3 and H4, indicating that at least these subunits of the nucleosome persist on such DNA. In fact, in some active genes the nucleosomes are displaced or "phased" in these regions (**Samal** *et. al.*, 1981).



In this experiment plants treated with compounds that interfere with DNA methylation produced a greatly increased number of flower stalks. Moreover, flowers that developed showed altered morphology. Control plants are shown in parts A, A and E whereas antisense – treated plants are shown in B, D and F.

(c) Modification of DNA sequences : DNA methylation. The genomic DNA of higher eukaryotes is modified following replication so that a large proportion of the cytosine (C) residues are present as **5-methylcytosine (5mC)**. However, such methylation has not been detected in the DNA of lower eukaryotes such as yeast and *Dictyosteltum*, nor in *Drosophila*. The percentage of methylated C residues in DNA relative to unmethylated C residues is highly variable, from less than 1 per cent in some insects to over 50 per cent in some higher plants and vertebrates. A much greater correlation exists between the methylation or under-methylation of sequence in the vicinity of gene promoters. For example, DNA of sperm is highly methylated, as in the DNA of the oocyte-specific 5S rRNA genes in adult tissues, whereas the sites around the coding regions of genes such as adult globin, ovalbumin, and immunoglobulin are under methylated in tissues in which they are expressed but are largely methylated in other cells in which they are not expressed.

There is evidence that the cytosine methylation in DNA alters the structure of the double helix in a fundamental way and favours the transition from B-form to Z-form DNA (**Bele** and **Felsenfeld**, 1981). It is possible that B–Z transition is itself involved in gene regulation and this may be the way in which DNA methylation has its effects on transcription.

(d) Modifications of histones. Histone component of chromatin is subject to three different post-synthetic modifications which have either direct or indirect effect on eukaryotic gene regulation : 1. **Histone methylation** affects only histones H3 and H4 and involves the irreversible methylation of a few lysine residues which alters the hydrophobic nature of the side chain of these histones. 2. **Histone phosphorylation** involves histone H1. Phosphorylation affects **serines** and **threonines**, changing them from a state of neutral charge to one of negative charge and is a reversible reaction. The state of phosphorylation of H1 protein varies through the eukaryotic cell cycle, and after H1 phosphorylation,

chromatin becomes much more strongly condensed, as it does in mitotic chromosomes. Evidently, activation of the **histone kinase** enzyme that is responsible for H1 phosphorylation may be the first step in the chain of events that leads to eventual chromatin condensation prior to mitotic cell division. 3. **Histone acetylation** is of two types. The first is the irreversible acetylation of the amino-terminal **serines** of histones H1, H2A and H4. These modifications seem to be associated with histone synthesis. The second is reversible acetylation of **lysine** residues in the amino terminal regions of histones H2A, H2B, H3, and H4. The acetylation converts the normally basic lysine side chain to a neutral acetyl lysine, and, thus, reduces the net basic charge of the amino-terminal ends of the affected histones. Both H3 and H4 can have up to four lysines in the acetyl form, and there is a strong correlation between this type of histone acetylation, especially **tetracetylation** of all available lysines, and transcriptionally active chromatin. Evidently, the acetylation of the core histone lysines would tend to loosen the nucleosomal structure which brings about the transition from a silent condensed gene to a transcriptionally active and extended one.

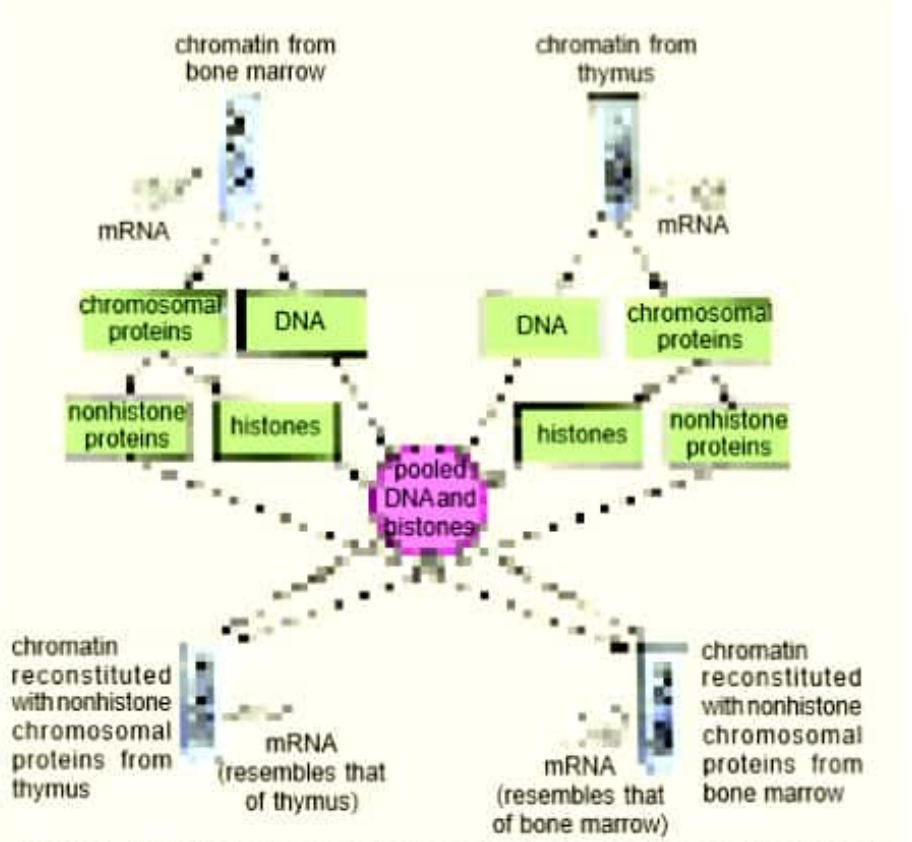


Fig. 8.9. Chromatin reconstitution experiment demonstrating the positive role of non-histone proteins in transcription.

(e) Transcriptional regulation by protein A24. The **A24** is an unusual hybrid protein, being a complex of histone **H2A** and the non-basic protein **ubiquitin**. The ubiquitin is covalently bound via the side chain amino group of lysine 19 of the histone. Some 10 per cent of H2A molecules are in the form of A24 and these specialized histones seem to be confined to interphase chromatin, disappearing as the chromosomes condense. A24 is found highly abundant in the chromatin of active genes.

(f) Gene regulatory molecules. Transcription of the eukaryotic genome is believed to be regulated by a variety of specific gene regulatory molecules which are produced by specific regulatory genes or by cytoplasm/cell surface. Examples of such gene regulatory molecules are the following :

(i) RNA polymerases. These enzymes are necessary for transcription and if they are short in supply, they tend to affect it. For example, there is a possible competition for type II polymerase (which is meant for hnRNA and mRNA) by the various promoter sequences that lie upstream of protein coding sequence.

(ii) Endonucleases. These enzymes are likely to affect the transcription, especially *in vitro* cell-free systems, by introducing into DNA nicks that may serve as initiation sites for some polymerases.

(iii) Topoisomerases, helicases and other DNA helix-destabilizing proteins. Various proteins are known that alter the three-dimensional structure of DNA and render it more available for processing which may affect transcription.

(iv) DNA methylase. This enzyme is likely to make DNA less available for transcription, and factors that antagonize methylation would enhance transcription.

(v) Histone acetylase and deacetylases. Such enzymes influence the rate of transcription by modulating acetylation.


Roy J. Britten

(vi) Factors such as ATP. Such molecules may influence the transcription rate by changing the available energy.

(vii) Ions and small molecules. Many ions such as those of **calcium, magnesium** and **manganese** directly affect chromatin conformation, which modulates gene activity.

Britten-Davidson model or gene-battery model of transcription regulation. This model of eukaryotic gene regulation at the level of transcription was proposed by **Britten** and **Davidson** in 1969 and later on elaborated by them in 1973. The gene-battery model assumes the presence of four classes of sequences : 1. **producer gene** which is comparable to a structural gene of prokaryotic operon; 2. **receptor site** is located adjacent to each producer gene and is comparable to operator gene of prokaryotic operon; 3. **integrator gene** which is comparable to regulator gene and is responsible for synthesis of an **activator RNA** that may or may not give rise to proteins before it activates the receptor site; 4. **sensor site** regulates the activity of integrator gene, which can be transcribed only when the sensor site is activated. The sensor sites are recognized by agents which, like hormones and proteins, change the pattern of gene expression. For instance, hormone-protein complex or a transcription factor may bind to a sensor site and cause the transcription of integrator.

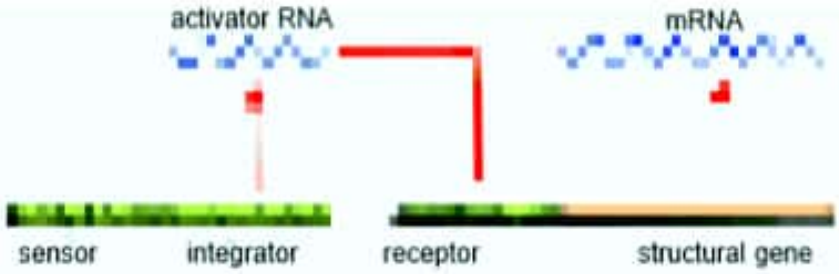


Fig. 8.10. Various components of Britten-Davidson's model for transcription regulation.

In this model, the genes (producer gene and integrator gene) are those sequences which are involved in RNA synthesis, while the receptor and sensor sites help only in recognition without taking part in RNA synthesis. Lastly, in Britten-Davidson's model, a set of structural genes controlled by one sensor site is called the **battery**.

C. Post-transcriptional Regulation

Steps that come between transcription and translation are described as **post-transcriptional** and are the following :

1. Some RNA is capped and tailed. The precise functions of capping and tailing of mRNA are not known, but they seem to serve to identify a message or potential message, and tailing may also help in the final export of this message from the nucleus.

2. RNA is processed to remove intron sequences. Introns removal and **splicing** together of the remaining exons during processing of hn RNA must be absolutely precise. This is in part engineered by a distinct group of nuclear particles (Sn RNPs containing U1, U2, U3, U4, U5 and U6 sn RNAs).

For example, **differential splicing** is used in different lymphocyte cells to produce different proteins from the same hn RNA molecule. As originally discovered by **Early** *et al.*, (1980), the two types of mRNA molecules are produced by part of an intron being omitted from one of the mRNAs but included in the exon splice used to produce the other mRNA. This allows production of two distinct proteins both immunoglobulins (Ig), but *one* with a long strand of hydrophobic amino acids at its carboxyl terminus, and the *other* with only a short length of relatively hydrophilic amino acids. The Ig molecule with the long hydrophobic peptide is membrane bounded within a lymphocyte, whilst the molecule with the terminal hydrophilic peptide is secreted from the cell. This change in splicing takes place within the life of a single lymphocyte cell and clearly explains the following observation. Immature lymphocytes retain antibody and simply insert the Ig molecules into their plasma membranes, whereas following stimulation with antigen, the same lymphocyte becomes secretory, releasing antibody molecules into circulation.

3. Most RNA is never exported from the nucleus. About 5 per cent of total transcribed RNA never leaves the nucleus. This is explained partly by removal of intron RNA and also by many RNA molecules which break up within the nucleus. The significance of this process is not clear, but some clues about the identification of RNA for export are coming to light. Although not all genes contain introns, most do, and it seems that the presence of some of these introns is essential for RNA export. In other words, introns are used as a means of identifying or ticketing the molecules that are to be passed out of the nucleus (see **Maclean** and **Hall**, 1987).

4. Message degradation rates are significant. The rate at which eukaryotic mRNA is degraded in the cytoplasm is highly variable. This implies that differential message breakdown is an important method of regulating not only the rate of gene expression, but also the lag between transcriptional shut-down and the cessation of specific translation. For example, the survival of histone mRNA during the cell cycle explains this fact very clearly. New histone is required in massive amounts immediately at the start of the S period of DNA synthesis to provide the new DNA with nucleosomes. Recently, it is discovered that the restricted availability of histone message is not achieved as a result of transcriptional control alone but by differential breakdown rates for histone message.

D. Translational Control

In bacteria, most mRNA molecules are translated about the same number of times with only fairly small variation from gene to gene. In eukaryotes translational regulation occurs in which a mRNA molecule is not translated at all until a signal is received. Translational control may involve the following mechanisms :

1. Extension of lifetime of the mRNA. An important example of translational regulation is that of **informosomes** or **masked mRNA**. Unfertilized eggs are biologically static, but shortly after fertilization many new proteins must be synthesized, for example, the proteins of the mitotic apparatus, the cell membranes, histones for nucleosome formation as well as others.

Localization of mRNA in the cytoplasm of a *Drosophila* egg.
Left – *bicoid* mRNA at anterior pole;
Right – *oskar* mRNA at posterior pole.

Unfertilized sea urchin eggs store large quantities of mRNA for many months in the form of mRNA-protein particles (= masked mRNA) made during formation of the egg. This mRNA is translationally inactive, but within minutes after fertilization, translation of these molecules begins. Here, the timing of translation is regulated; the mechanism for stabilizing the mRNA, for protecting it against RNases, and for activation are still unknown.

2. Regulation of rate of protein synthesis. This type of regulation also occurs in mature unfertilized eggs. These cells need to maintain themselves but do not have to grow or undergo a change of state. Thus, the rate of protein synthesis in eggs is generally low. This is not due to inadequate supply of mRNA but of a limitation of an as-yet-unidentified element, called the **recruitment factor** which apparently interferes with formation of the ribosome-mRNA complex. A good example of translational control is the extension of the lifetime of silk fibroin mRNA in the silkworm *Bombyx mori*. During cocoon formation the silk gland of the silkworm predominantly synthesizes a single type of protein, **silk fibroin**. Since the worm takes several days to construct its cocoon, it is the total amount and not the rate of fibroin synthesis that must be great; the silk worm achieves this by synthesizing a fibroin mRNA molecule that is very long lived.

Transcription of the fibroin gene is initiated at a strong promoter by an unknown signal and about 10^4 fibroin mRNA molecules are made in a period of several days (such a synthesis forms an example of transcriptional regulation). A typical eukaryotic mRNA molecule has a life-time of about three hours before it is degraded. However, the fibroin mRNA survives for several days during which each mRNA is translated repeatedly to yield 10^5 fibroin molecules. Thus, each gene is responsible for the synthesis of 10^9 protein molecules in four days. Altogether the silk gland makes 300 µg or 10^{15} molecules of fibroin during this period. If the lifetime of mRNA were not extended, either 25 times as many genes would be needed or synthesis of the required fibroin would take about 100 days.

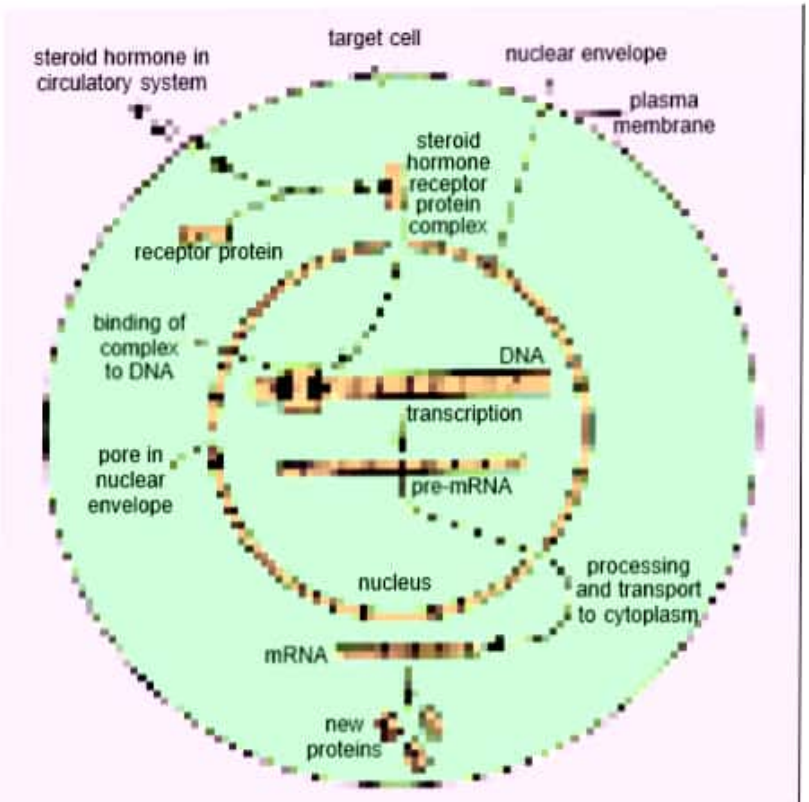


Fig.8.11. Mode of action of steroid hormones on the gene expression (*i.e.*, transcription) (after Gardner, *et al.*, 1991).

E. Post-translational Modification of Proteins to Make Them Active Ones

Some proteins are altered after synthesis, usually by partial degradation or trimming, as for example, by the enzymatic removal of the central section of

the **proinsulin** molecule to yield the active protein, **insulin**. For their activity many proteins also depend on being complexed into compound proteins together with other subunits, either the same or different in nature. Such post-translational control mechanisms do play a significant role in determining the activities of differential cells. For example, **haemoglobin** production is highly dependent on the availability of haem to complex with globin protein subunits and may be deficient in cases of iron-dependent anaemia.

HORMONAL CONTROL OF GENE EXPRESSION

In higher plants and animals, intercellular communication is a very important phenomenon. Signals originating in various glands and/or secretory cells somehow stimulate **target tissue** or **target cells** to undergo dramatic changes in their metabolic patterns. These changes frequently include altered pattern of differentiation that are generally dependent on altered patterns of gene expression. **Peptide hormones** such as insulin, epinephrine, etc., and **steroid hormones** such as estrogen, progesterone, testosterone (in higher animals, *e.g.*, mammals) and ecdysone (in insects). In higher animals, hormones are synthesized in various specialized secretory cells (*i.e.*, endocrine cells) and are released into the blood stream. The peptide hormones do not normally enter cells because of their relative large size. Their effects appear to be mediated by receptor proteins located in target-cell membranes and by the intracellular levels of **cyclic AMP (cAMP)** (called **secondary messenger**). The cAMP activates a protein kinase (e.g., A-kinase) which phosphorylates (activates) many specific enzymes. The steroid hormones, on the other hand, are small molecules that readily enter cells through the plasma membrane. Once inside the appropriate target cells, the steroid hormones become tightly bound to **specific receptor proteins** which are present only in the cytoplasm of target cells. The hormone-receptor protein complexes activate the transcription of specific genes are sets of genes according to following two methods (= hypotheses): 1. The hormone receptor protein complexes interact with specific non-histone chromosomal protein and this interaction stimulates the transcription of the correct genes (**J. Stein**, **G. Stein** and **L. Kleinsmith**, 1975). 2. The hormone receptor protein complexes activate transcription of target genes by binding to specific DNA sequences present in the *cis*-acting regulatory regions (the enhancers and promoter regions) of the genes (**R.M. Evans**, 1988). In both of these cases, these hormone- receptor protein complexes would function as positive regulators (or "activators") of transcription much like the CAP-cAMP complexes in prokaryotes.

During development of dipteran flies (*Drosophila melanogaster* and *Chironomus tentans*) the steroid hormone **ecdysone** is released and triggers moulting. If larvae of these insects are treated with ecdysone at stages of development prior to or between moultings, patterns of chromosome puffing occur that are identical to those occurring during natural moultings. Ecdysone tends to affect the gene expression at the level of transcription.

REVISION QUESTIONS AND ANSWERS

1. Describe the operon model for regulation of gene activity.