（内部交流）

ADVANCES IN GENETICS

VOLUME 18

Edited by

E. W. CASPARI

Department of Biology
University of Rochester
Rochester, New York

1976

ACADEMIC PRESS • NEW YORK SAN FRANCISCO LONDON

A Subsidiary of Harcourt Brace Jovanovich, Publishers

（内部交流）

ADVANCES IN GENETICS

VOLUME 18

Edited by

E. W. CASPARI

Department of Biology
University of Rochester
Rochester, New York

1976

ACADEMIC PRESS • NEW YORK SAN FRANCISCO LONDON

A Subsidiary of Harcourt Brace Jovanovich, Publishers

ACADEMIC PRESS, INC.
111 Fifth Avenue, New York, New York 10003

United Kingdom Edition published by
ACADEMIC PRESS, INC. (LONDON) LTD.
24/28 Oval Road, London NW1

LIBRARY OF CONGRESS CATALOG CARD NUMBER: 47-30313

ISBN 0-12-017618-1

PRINTED IN THE UNITED STATES OF AMERICA

CONTENTS

Developmental Genetics of Homoeosis

WILLEM J. OUWENEEL

Drosophila Gene–Enzyme Systems

JAMES B. COURTRIGHT

Genetics, Cytogenetics, and Evolution of Mosquitoes

JAMES B. KITZMILLER

REGULATION OF GENE EXPRESSION IN PROKARYOTIC ORGANISMS

Robert F. Goldberger, Roger G. Deeley, and Kathleen P. Mullinix

Laboratory of Biochemistry, National Cancer Institute,
National Institutes of Health, Bethesda, Maryland

I. Introduction

One of the most striking characteristics of living systems is that they function in an orderly manner despite their high degree of complexity. One workable definition of regulation, in fact, is that it is the set of mechanisms that allows organisms to maintain this orderly functioning. It is important to realize, however, that regulation was not superimposed upon living systems; orderly processes are simply more successful than are disorderly ones, and therefore tend to be preserved through the evolution-

ary process by conferring advantages upon organisms that possess them. The thousands of chemical reactions occurring in cells are controlled by regulatory mechanisms that operate at many different levels. In this chapter we will concentrate on regulation at the level of gene expression.

The work of Anfinsen (1973) and his colleagues has demonstrated that in order for a gene to specify the complete and final structure of a native protein, it need do no more than specify the amino acid sequence of that protein. Acting as *structural genes*, which carry the specifications for the amino acid sequences of proteins is not, however, the only function of DNA in the cell. There are also genes for transfer RNA and for ribosomal RNA; there are regions of DNA that serve as signals for the initiation of RNA synthesis—promoters—and for termination of RNA synthesis. In addition, there are genes of another type—control genes—that regulate the frequency with which neighboring structural genes are transcribed into messenger-RNA (mRNA), thereby providing the cell with a means for regulating the intracellular concentrations of specific proteins. Many structural genes, perhaps the vast majority, are not regulated in this way. They are said to be *constitutive;* the intracellular concentrations of the proteins they specify remain essentially constant under a wide variety of conditions.

Those structural genes for which the transcription frequency can be regulated have been the subject of intensive investigation over the past two decades. In a fairly high proportion of cases these genes have been found to be clustered on the bacterial chromosome with one or more other structural genes that specify functionally related proteins. A cluster of functionally related genes is known as an *operon;* the proteins encoded in the genes of a given operon are ordinarily the enzymes that catalyze the several steps of a metabolic pathway. For example, all ten enzymes that catalyze the conversion of ATP and phosphoribosyl pyrophosphate to the amino acid histidine are encoded in the contiguous genes of the histidine operon (Ames and Hartman, 1963); the same organization is found for the three enzymes that catalyze the degradation of galactose (Buttin, 1963a,b), the six enzymes that catalyze the biosynthesis of tryptophan (Yanofsky, 1971), and the enzymes of many other metabolic pathways.

One of the important features of operons is that they are transcribed into messenger-RNA at a frequency that varies in response to changes in the environment of the bacterial cell. Each operon is transcribed into a single, polycistronic* mRNA (Martin, 1963); for this reason, when the

*The term *polycistronic* is used to indicate a single molecule of mRNA that carries the specifications for more than one polypeptide chain, each polypeptide chain being specified by one cistron.

frequency of transcription of an operon increases or decreases, the intracellular levels of *all* the enzymes encoded in the genes of that operon increase or decrease, respectively. The polycistronic nature of the mRNA transcribed from each operon is a reflection of the fact that each operon has only one primary promoter site at which RNA polymerase binds prior to initiating transcription. Although other promoters have been identified within certain operons, these are always secondary promoter sites for which the affinity of the polymerase is relatively low in comparison with the primary promoter at the beginning of the operon; therefore, they do not ordinarily function to a significant degree. Thus, since RNA polymerase binds at a single site for each operon, normally only one additional region, a regulatory region, located between the promoter and the structural gene(s), would be required to provide a mechanism by which the cell could regulate the frequency with which polymerase molecules are permitted to transcribe the operon once they have become bound. This regulatory region, which serves to control the frequency of transcription for each operon, functions much as does a valve, allowing greater or smaller numbers of polymerase molecules to traverse the operon per unit time, depending upon whether it is more "open" or more "closed." The frequency with which an operon is transcribed thus depends not only upon the affinity of its promoter for RNA polymerase, but also upon the degree to which the regulatory region restricts the passage of polymerase molecules from the promoter into the structural genes.

The regulatory regions are of two types. One of these, the *operator* gene, is a valve that by itself is "open," allowing polymerase molecules through at a relatively high frequency; it is regulated by being progressively closed. The other, the *positive activator* gene, is a valve that by itself is "closed," restricting the passage of polymerase molecules to a low frequency; it is regulated by being progressively opened. In both cases the mechanism for altering the frequency with which polymerase is able to transcribe the operon involves a specific protein and a specific small molecule for each operon. The regulatory protein for the operator gene is known as a *repressor*. This protein binds specifically to the operator, closing it to the passage of polymerase molecules, thereby diminishing transcription of the operon. For the positive activator gene the regulatory protein is known as an *activator*. This protein binds specifically to the positive activator gene, opening it to the passage of polymerase molecules, thereby increasing transcription of the operon. In both cases, the specificity resides in the recognition of the DNA of the regulatory region by the regulatory protein.

The small molecules involved in altering the frequency of operon transcription interact with and affect the properties of the specific regulatory

proteins. In the case of repressible operons, the regulatory protein has no activity by itself—it is called an *aporepressor*. It takes on the properties of a repressor only when it binds the specific small molecule, known as the *corepressor*. In general, the small molecule in a system of this type is the end product of a biosynthetic pathway or a molecule closely related to the end product. The pathway is, of course, the one catalyzed by the enzymes encoded in the regulated operon. In the case of inducible operons, the regulatory protein is, by itself, the active repressor. The small molecule, known as the *inducer*, binds to the repressor and thereby either renders it unable to bind to the operator gene (negative control) or imparts to it a new activity—namely, the ability to bind to the positive activator gene (positive control). In general, the small molecule in such systems is the substrate of a catabolic pathway or a molecule closely related to the substrate. The pathway is the one catalyzed by the enzymes encoded in the regulated operon.

The differences between inducible and repressible systems discussed above are well suited to the different roles they play in regulating the metabolism of bacterial cells. The major evolutionary advantage of inducible systems is that they allow the organism to utilize energy-rich substrates not usually present in the environment, while sparing the organism the waste of manufacturing the enzymes necessary to metabolize rare substrates when a more common one is available. The lactose operon of *Escherichia coli* may serve as an example. When glucose is available in the medium, the organism utilizes this sugar, and the enzymes for uptake and catabolism of lactose are repressed. When glucose is depleted from the medium, however, and lactose is present, then lactose enters the cell, is converted in one step to allolactose, and this compound acts as inducer of the lactose operon (Jobe and Bourgeois, 1972). Allolactose binds to the specific repressor, removing it from the operator gene of the lactose operon, thereby causing a greatly increased frequency of transcription of the operon. The intracellular levels of the proteins involved in lactose degradation rise dramatically, and the organism thrives in its new environment, utilizing lactose as its source of energy. Thus, as in this example, inducible systems are ordinarily catabolic and function in an adaptive capacity.

In contrast to inducible systems, repressible systems ordinarily involve biosynthetic pathways and function in an economic capacity. For example, the pathway for the synthesis of the amino acid tryptophan is among the systems regulated by repression (for review, see Yanofsky, 1971). The enzymes that catalyze the six steps of this pathway in *E. coli* are encoded in the genes of a single operon, controlled by one operator gene. The repressor, as is the rule for repressible systems, is composed of a

protein, the aporepressor, and a small molecule, the corepressor. The corepressor in this case is the end product of the biosynthetic pathway, tryptophan. Thus, when the organism grows in the presence of exogenous tryptophan, the intracellular level of this amino acid is sufficiently high to maintain repression of the tryptophan operon; the enzymes for tryptophan biosynthesis are made at a relatively low basal rate, and the pathway functions at a low level. When, however, the organism grows in the absence of exogenous tryptophan, the amount of this amino acid that is synthesized endogenously at the basal rate is not sufficient to saturate the aporepressor. Therefore the operon becomes derepressed; the enzymes for tryptophan biosynthesis are produced at a higher rate, and the rate of tryptophan synthesis consequently increases. If tryptophan is added back to the derepressed culture, the tryptophan operon becomes repressed once more. Thus, the rate at which the enzymes for tryptophan biosynthesis are manufactured is closely geared to the availability of tryptophan, allowing the organism to conserve the energy that would be wasted in synthesizing these enzymes when they are not needed.

The fact that repressible systems arose and were maintained through the evolutionary process has been explained by the hypothesis that organisms that possess regulated systems are at a selective advantage over organisms that do not. This hypothesis was tested experimentally by Zamenhof and Eichhorn (1967). They inoculated a culture with an equal number of cells of two different strains of *E. coli*. These strains differed in only one respect—the ability to regulate tryptophan biosynthesis. One strain was mutated so that it was constitutive; it had lost the ability to repress the tryptophan operon. The other strain was repressible; its tryptophan operon became repressed whenever tryptophan was available exogenously. The mixed culture of the two strains was grown in liquid medium containing tryptophan. The cells of the first strain continued to make the enzymes for tryptophan biosynthesis at a high rate even though the amino acid was available in its external medium, whereas the second strain produced these enzymes at the lowest possible (basal) rate. At the end of a few days the normally regulated strain was found to have overgrown the constitutive strain, taking over essentially full possession of the culture. Evidently, even the relatively small amount of energy saved by repressing a single operon under the appropriate conditions is sufficient to be the crucial factor in survival of an organism.

Having discussed some of the general principles involved in how bacterial cells modify expression of their genetic potential, we will next present several specific examples of regulated systems. These systems were chosen to illustrate the basic mechanisms that constitute the form in which the model of Jacob and Monod (1961) in currently understood.

They include both positive and negative control of gene expression; induction and repression of specific protein synthesis; autogenous regulation, a mechanism by which a protein regulates expression of its own structural gene; catabolite repression, a mechanism by which the bacterial cell is able to discriminate among a variety of energy sources; and stringency, a mechanism by which the cell coordinates the biosynthesis of macromolecules of different types. We will then consider RNA polymerase, an enzyme inevitably involved in all gene expression, discussing the structure and function of this enzyme, as well as the mechanisms by which its activity and specificity may be altered to modify gene expression. Finally, we will discuss regulation of specific protein synthesis through control of translation of mRNA, a problem less clearly understood in terms of its mechanism and prevalence in bacterial cells.

II. Examples of Regulated Systems

In this section we shall consider, in some detail, three operons. Two of them, the lactose and tryptophan operons, have already been mentioned in the introduction. They are probably the most well understood examples of an inducible and a repressible system, respectively. The third, the arabinose operon, is a positively controlled system—a system that involves a regulatory mechanism not specifically predicted by the model proposed by Jacob and Monod (1961).

These three systems have been chosen to illustrate some of the basic mechanisms involved in regulation at the level of operon expression. Several variations of, and elaborations on, these mechanisms are known to exist in other systems, but rather than consider these other systems in detail, we shall merely draw attention to particular features of their regulation.

A. The Lactose Operon

1. General

Studies on the system for lactose utilization in *E. coli* have been in progress since the 1940s (Monod and Audureau, 1946). The contribution that these studies have made to our understanding of the basic mechanisms of genetic control has indeed been enormous. The explanation given by Jacob and Monod (1970) for choosing to study *lac,* rather than one of the other adaptive enzyme systems then under study at the Pasteur Institute, may seem almost trivial in retrospect. The reason was the possi-

bility of synthesizing nonmetabolizable galactoside analogs, capable of inducing production of the *lac* system enzymes (gratuitous inducers). Using such inducers, they were able to dissociate enzyme induction from enzyme activity, thereby disproving earlier theories that had suggested that interaction of a substrate with the active site of an inducible enzyme was an integral part of the induction process.

Two timely advances in microbial genetics facilitated the subsequent genetic and physiological studies on the lactose system. The first was an understanding of the mechanism of chromosome transfer from male to female cell (Wollman and Jacob, 1955). This understanding permitted an early genetic analysis of the *lac* system and also the construction of strains that temporarily contained two copies of the *lac* genes (temporary merozygotes). The second was the isolation of an episome, F'*lac* (Jacob and Adelberg, 1959) that carried the genes of the *lac* operon. This allowed the construction of stable bacterial strains that were diploid for the genes of the *lac* operon (stable merozygotes). It was then possible to study not only the effects of a mutation on expression of genes on the same piece of DNA (cis effects), but also the effects of a mutation on expression of genes to which the mutated segment was not covalently linked (trans effects). This technique allowed the demonstration that the regulator gene (*lac i*) exerted its control by means of a freely diffusible product (the repressor), which probably interacted with a site on DNA (the operator) that remained covalently linked to the structural genes of the *lac* operon (Jacob and Monod, 1961).

A third technique that should be mentioned here was the development of a method for obtaining a variety of *E. coli* genes on the transducing bacteriophages, $\phi 80$ and λ. Beckwith and Signer (1966) isolated both $\phi 80$ and λ carrying the lactose operon in place of some of their own genes. The technique greatly simplified the construction of bacterial strains that were diploid for the lactose operon and provided a source of DNA which, when compared with bacterial DNA, is enriched approximately 100 times for the bacterial genes that it carries. The availability of bacterial genes on bacteriophage DNA has been extremely useful in the development of *in vitro* systems that have been used to study the regulation of a variety of bacterial operons.

The lactose operon (see Fig. 1) consists of three contiguous structural genes, *z*, *y*, and *a*, which are transcribed in that order. The *z* gene specifies the enzyme, β-galactosidase, which hydrolyzes lactose to glucose and galactose. The *y* gene product is galactoside permease, which is involved in the transport of lactose into the cell and its subsequent accumulation. The enzyme thiogalactoside transacetylase is specified by the *a* gene, and as yet the physiological function of this enzyme is not known.

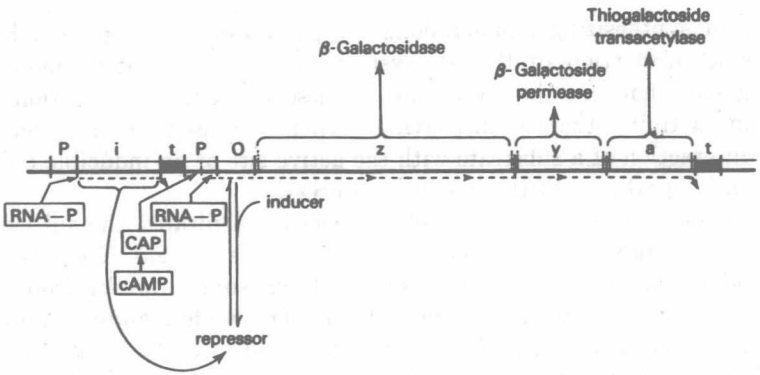

FIG. 1. A model of the lactose operon of *Escherichia coli* together with the closely linked regulatory (*i*) gene. The three structural genes of the operon (*z*, *y*, and *a*) are transcribed in the direction indicated by the dashed arrow. Abbreviations: P, promoter; O, operator; *t*, terminator; RNA-P, RNA polymerase; CAP, catabolite gene activator protein; cAMP, adenosine-3′:5′-cyclic-monophosphate.

The production of these three enzymes is coordinately induced by lactose or any of a variety of gratuitous inducers. The fully induced rate of enzyme synthesis is approximately 1000 times greater than the basal, or repressed, rate. The three structural genes are under the control of a single operator that is adjacent to the *z* gene. Mutations in the operator that reduce its affinity for repressor cause an increase in the rate of enzyme synthesis in the absence of inducer (O^c mutations). These increases may vary from a fewfold over that of the uninduced level to levels approaching those found under conditions of full induction (Sadler and Smith, 1971). Adjacent to the operator and distal from the *z* gene is the promoter. Genetically, this site has been defined by mutations that alter the maximal rate of expression of the operon, either increasing it or decreasing it. Conceptually, the promoter was originally considered to be the site of initiation of transcription, and probably the site to which RNA polymerase bound. It is now apparent that the genetically defined *lac* promoter is also involved in the mechanism of catabolite repression (Section IV).

The gene specifying the repressor (the *i* gene) of the *lac* operon is adjacent to the *lac* promoter and operator region. It is transcribed in the same direction as is the operon (Kumar and Szybalski, 1969), but apparently RNA polymerase molecules that initiate transcription at the *i* gene promoter do not read through into the *lac* operon. It is postulated that a transcriptional termination signal exists between the *i* gene and the *lac* promoter (Miller *et al.*, 1968b). The gene appears to be expressed

at a low constitutive rate. The repressor, which is a tetramer of identical subunits, normally constitutes approximately 0.002% of the cell's protein (10–20 molecules of repressor per cell).

Induction of the operon by its apparent physiological inducer, lactose, will occur only if the enzyme, β-galactosidase, is functional. This is because, in order to cause induction, lactose must first be converted to allolactose, the true inducer of the operon (Jobe and Bourgeois, 1972). Furthermore, unless galactoside permease is active, the level of lactose in the cell rapidly falls below that required for continued induction. Both these complications can be avoided in studies of the induction process by using one of the gratuitous inducers of the operon, such as isopropylthiogalactoside (IPTG).

If IPTG is added to an exponentially growing culture of *E. coli*, the differential rate of β-galactosidase synthesis begins to increase after a lag of 3–4 minutes (Pardee and Prestidge, 1961). Even if the cells are exposed to inducer for only 20 seconds, however, a pulse of enzyme synthesis occurs a few minutes later. This pulse of enzyme synthesis is concomitant with the completion of transcription of the β-galactosidase gene by RNA polymerase molecules that began transcription of the operon during the brief exposure of the cells to IPTG. By measuring the decay in the rate of β-galactosidase synthesis following a brief exposure to the inducer, it was possible to determine the half-life of the β-galactosidase message. The half-life obtained in this way was found to be 1 minute. This short half-life of the message fulfilled an important prediction of the model of genetic control in which regulation occurs at the level of transcription (Jacob and Monod, 1961).

In the following discussion we will consider, in greater detail, some of the biochemical and genetic studies that have led to our present understanding of the structure and function of the individual regulatory elements of the operon.

2. The Repressor

As mentioned previously, the *lac* repressor normally constitutes only 0.002% of the cell's protein. By a series of genetic manipulations, however, a mutant strain of *E. coli* has been selected that has a *lac i* gene promoter (i^{sQ}) 50 times more efficient than the normal promoter. It has been possible to raise the repressor levels in the cell still further by infecting an *E. coli* with a temperature-inducible, defective bacteriophage that carries the *lac* region with the i^{sQ} mutation. The induction of a strain of *E. coli* lysogenic for this bacteriophage produces multiple copies of the i^{sQ} gene and raises the level of repressor to 1–2% of the soluble cell protein (Gilbert and Müller-Hill, 1970). Thus, it has been possible to obtain the

repressor in quantities large enough to purify it to homogeneity. Many of the properties of this protein have now been studied, and its complete amino acid sequence is known (Beyreuther *et al.*, 1973).

The binding of *lac* repressor to double-stranded DNA of the *lac* operator has been demonstrated *in vitro* (Riggs *et al.*, 1968). This binding is extremely tight, the dissociation constant being 10^{-13} to 10^{-14} *M*. One implication of the tightness of the binding is that, once formed, the operator–repressor complex is very slow to dissociate; if the rate of formation is assumed to be diffusion-limited, then a decay time of 10^4 seconds is predicted. An estimate of the decay times of the complexes formed *in vitro* between the *lac* operator and both normal and mutated *lac* repressors of several types has been made by Jobe *et al.* (1972). The half-time for decay of the normal repressor–operator complex is 30 minutes. Obviously, if the inducer prevented repression by a simple competition for the operator binding site of the repressor, induction would be limited by the rate at which the complex decayed. The fact that induction occurs much more rapidly than this indicates that the inducer must be able to interact with the operator–repressor complex and destabilize it. Riggs *et al.* (1970a) have demonstrated *in vitro* that this prediction is correct. Estimates of the affinity of the *lac* repressor–inducer complex for the *lac* operator *in vivo* and *in vitro* suggest an affinity 10^3 to 10^4 times weaker than that of the repressor alone (Jobe and Bourgeois, 1973; Gilbert and Müller-Hill, 1970). The intracellular concentration of repressor is such that in the absence of inducer the operon should be 99.9% shut down and in the presence of saturating amounts of inducer only 5% shut down. These estimates may be in error since they assume that the total concentration of repressor is identical with its "free" concentration. The repressor, however, has a significant affinity for DNA other than the *lac* operator (this affinity increases with the A-T content of the DNA, Lin and Riggs, 1970). Therefore, some proportion of it may be bound to nonoperator DNA. The affinity of the repressor for DNA in general may fulfill a useful function. One might expect the rate of formation of the repressor–operator complex to be limited by the rate at which the repressor can diffuse to the operator. An estimated rate, taking the size of the operator and repressor into consideration, would be 1 to 2×10^9/mole · second. The measured rate, however, is 5×10^9/mole · second. This suggests that the target for the repressor is considerably larger than any estimate of the size of the *lac* operator. A model has been proposed in which the repressor may diffuse to a large site on DNA, then rapidly move one-dimensionally along the DNA to the operator region (Riggs *et al.*, 1970b).

Although the repressor has a considerable affinity for DNA from a variety of sources, as well as synthetic polynucleotides [dissociation constant

10^{-8} *M* for poly(dAT)], the binding of repressor to operator is at least five orders of magnitude tighter than the nonspecific binding. Thus, it has been possible to demonstrate the specificity of the repressor–operator interaction *in vitro* (Lin and Riggs, 1972).

3. The Operator

In the model proposed by Jacob and Monod (1961), the operator was considered to be the binding site for the repressor and to serve as the site for initiation of transcription. We have already considered some of the evidence that demonstrates that the operator is indeed the binding site of the repressor. Whether or not it is the site of initiation of transcription is less clear.

The original proposal that the operator might be the site of initiation of transcription was based upon the existence of mutations, thought to be in the operator region, that abolished operon expression (0° mutations). However, it was subsequently shown that these mutations were in the *z* gene (Beckwith, 1964) and that they exerted their effect at the translational level by introducing a nonsense codon in the *z* gene (a codon that specifies termination of polypeptide chain synthesis). Furthermore, identification and localization of the promoter region adjacent to the operator and distal from the *z* gene (Ippen *et al.*, 1968) led to the assumption that the operator was transcribed but was probably not the site of initiation of transcription.

It was not possible to determine conclusively whether or not the operator is completely distinct from the *z* gene until very recently, although it has been known for some time that deletion of most or all of the *lac* operator does not affect the N-terminal sequence of β-galactosidase produced in the deletion-carrying strain (Reznikoff and Beckwith, 1969; Bhorjee *et al.*, 1969). However, the possibility that the operator was translated into a polypeptide subsequently cleaved from functional β-galactosidase could not be rigorously excluded on the basis of these data. Recently, Maizels (1973) has sequenced the RNA synthesized *in vitro* from a DNA fragment obtained from a *lac* transducing bacteriophage. This fragment includes the *lac* operator and the N-terminal portion of the *z* gene. It has been possible to locate the translational initiation signal at the beginning of the *z* gene from a knowledge of the N-terminal sequence of β-galactosidase. There appears to be a sequence of 17 or 18 nucleotides between the end of the operator and the translational initiation codon for β-galactosidase synthesis. Thus, the *lac* operator is completely distinct from its adjacent structural gene.

The shortest sequence of nucleotides that would be expected to occur only once in a polynucleotide the size of the *E. coli* genome was pre-

dicted, on statistical grounds, to be twelve (Gilbert and Müller-Hill, 1967). Thus, the *lac* operator–repressor interaction is expected to involve at least 12 base pairs. An extensive genetic analysis of the *lac* operator (Sadler and Smith, 1971) suggests that at least 12, and probably 16, distinct point mutations can reduce operator–repressor binding. These mutations can be divided into 8 classes on the basis of the extent to which they alter operon expression. Each of these classes consists of mutations at two sites in the operator region, the arrangement of which suggests that the operator is bilaterally symmetrical.

Walter Gilbert and his colleagues (Gilbert, 1971) have estimated the size of the *lac* operator by an entirely different technique, involving isolation of the portion of DNA that is protected by *lac* repressor against enzymatic degradation. These workers started with a bacteriophage carrying the lactose operon in place of some of its own genes. They fragmented the DNA of this phage by sonication, obtaining fragments approximately 1000 nucleotides in length. After addition of purified repressor to the fragments, the mixture was digested with deoxyribonuclease. The DNA was all degraded except for the specific portion that was protected by having repressor bound to it. After separation of this DNA–repressor complex from the reaction mixture, the repressor was removed from the small DNA fragment by addition of the gratuitous inducer, IPTG. The fragment obtained in this way was found to have 24 base pairs, with a length approximately the same as the diameter of the repressor. Although the DNA fulfills the cardinal criterion expected for the *lac* operator—specific binding of *lac* repressor, reversible specifically by inducer—it is not possible to specify whether the *lac* operator is actually longer than that part which is protected from enzymatic degradation by having repressor bound to it. The design of the experiment was such that one would not expect to obtain a fragment of DNA larger than the diameter of the repressor. RNA polymerase has been used to transcribe the protected fragment and the RNA produced has been sequenced (Gilbert and Maxam, 1973). The sequence of the operator derived from that of the RNA synthesized *in vitro* is shown in Fig. 2. It can be seen that

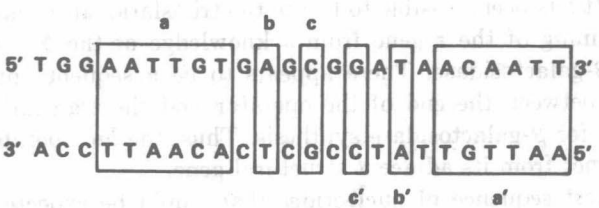

FIG. 2. Nucleotide sequence of the *lac* operator region (Gilbert and Maxam, 1973).