Regulatory Biology

Edited by James C. Copeland
and George A. Marzluf

EDITED BY JAMES C. COPELAND AND

GEORGE A. MARZLUF

Regulatory Biology

OHIO STATE UNIVERSITY PRESS: COLUMBUS

Manufactured in the United States of America

Library of Congress Cataloguing in Publication Data

Copeland, J C
Regulatory biology.

(Ohio State University biosciences colloquia)
Includes index.
1. Cellular control mechanisms. I. Marzluf, G. A., joint author. II. Title. III. Series: Ohio. State University, Columbus. Ohio State University biosciences colloquia.
QH604.C66 574.8'76 77-2256
ISBN 0-8142-0262-4

Preface

Regulatory biology is one of the most central and important topics of modern biological investigation—and rightly so, in view of the presence and crucial role of regulatory events in all living organisms. Control mechanisms can be detected in the simplest organisms known, the small single-stranded RNA phage, which possess only enough genetic material to encode three proteins. At the other extreme of the biological spectrum, the complex process of differentiation and development of vertebrates appears to be primarily dependent upon regulatory phenomena, particularly selective gene expression. It now appears that all of the significant activities of living cells are subject to regulation, which explains the fascination and importance of regulatory biology.

The objective of the Ohio State University Biosciences Colloquium on Regulatory Biology was to examine a variety of regulatory phenomena, which control widely divergent cellular activities. To provide perspective and to illustrate the unity of regulatory principles, the presentations ranged from control of transcription and DNA replication in coli phage lambda to development in *Dictyostelium* and in amphibians. Between these far-spaced limits, other equally exciting and informative control processes were analyzed, namely, the lac operon, guanosine tetraphosphate control, autogenous control of gene expression, the cell cycle, chromosomal replication, and fungal gene expression. The presentation of contributed papers on related topics served very well to strengthen and broaden the inquiry into, and our appreciation of, regulatory biology.

The colloquium was attended by more than three hundred persons from all parts of the United States and from Canada. We were most pleased by the enthusiastic response. We are grateful to Dean Richard Böhning and Associate Dean Richard Moore of the College of Biological Sciences for their support and encouragement. Our appreciation also goes to the speakers, the Colloquium Series Committee, our students, and to the many others who contributed to the colloquium in so many ways.

Second Annual Biosciences Colloquium
College of Biological Sciences
Ohio State University
4–6 September 1975

REGULATORY BIOLOGY

Organizers

James C. Copeland, Associate Professor of Microbiology, Ohio State University

George A. Marzluf, Professor of Biochemistry, Ohio State University

Speakers

Suzanne Bourgeois, Salk Institute for Biological Studies, San Diego

Michael Cashel, Section on Regulatory Cell Physiology, Laboratory of Molecular Genetics, National Institute of Child Health and Human Development, National Institutes of Health

James C. Copeland, Department of Microbiology, Ohio State University

Robert Dottin, Department of Biology, Johns Hopkins University

Robert F. Goldberger, Laboratory of Biochemistry, National Cancer Institute, National Institutes of Health

Thomas Humphreys, Department of Biochemistry, Kewalo Marine Laboratory, University of Hawaii

George A. Marzluf, Department of Biochemistry, Ohio State University

Waclaw Szybalski, McArdle Laboratory for Cancer Research, University of Wisconsin

Erik Zeuthen, Biological Institute, Carlsberg Foundation, Copenhagen

Jonathan King, Department of Biology, Massachusetts Institute of Technology (paper does not appear in this volume)

Contents

Regulatory Biology

WACLAW SZYBALSKI

Initiation and Regulation of Transcription and DNA Replication in Coliphage Lambda

1

INTRODUCTION

It has been said that the decades of the 1950s and 1960s were the age of simplification. I cannot recall whether this statement pertained to science or politics, but there is little doubt in my mind that in the complex field of regulatory biology the early, very basic concepts of the operon and its control elements, the promoter and operator, were actually simplifications. In this contribution, an updated and expanded summary of my previous reviews, I shall outline the intricate network of transcriptional controls active during development of the *Escherichia coli* bacteriophage λ and discuss current models for the structure and function of the promoters, operators, terminators, the origin of DNA replication, and other regulatory recognition sites. It will be apparent that the age of simplification is over since, even for as simple a virus as λ, the developmental controls are quite complex.

The references in this review will be mainly to studies published after 1973; earlier references are contained in previous reviews from our laboratory (Szybalski, 1969, 1970, 1971, 1972, 1974a,b,c; Szybalski et al., 1969, 1970) and others (Dove, 1968; Echols, 1971a,b, 1972; Eisen and Ptashne, 1971; Herskowitz, 1974; Kourilsky, Bourguignon, and Gros, 1971; Maniatis et al., 1975; Ptashne, 1971; Ptashne et al., 1976).

BACTERIOPHAGE λ

Coliphage λ is a conventional phage of medium size; the head contains a linear DNA molecule of 31.8×10^6 daltons, corresponding to 48,000 nucleo-

McArdle Laboratory for Cancer Research, School of Medicine, University of Wisconsin, Madison, Wisconsin 53706.

tide pairs (Davidson and Szybalski, 1971). Upon phage infection λ DNA is injected into the *Escherichia coli* host cell through the flexible phage tail. After entering the bacterial cell, the linear DNA molecule is converted into a circular form (fig. 1a,b) by the covalent joining of its two single-stranded cohesive ends, each composed of 12 complementary deoxynucleotides, employing the DNA ligase of the host.

Depending on the phage strain and the physiological state of the bacterial host, λ infection may evoke either a lytic or a lysogenic response. In the *lytic* response λ functions become sequentially expressed, λ DNA replicates, heads and tails are synthesized, and ultimately a new crop of mature phage progenies is produced, culminating in death and lysis of the host cell. In the *lysogenic* response the lytic cycle is interrupted, the host cell survives, and the circular λ genome is linearized while inserted into the bacterial genome, as depicted in figure 1 (c + d = e). This is possible because the Int enzyme, a λ gene product, mediates the insertion of the phage genome, and because λ can elicit synthesis of the *repressor,* the product of gene *c*I, which blocks all the

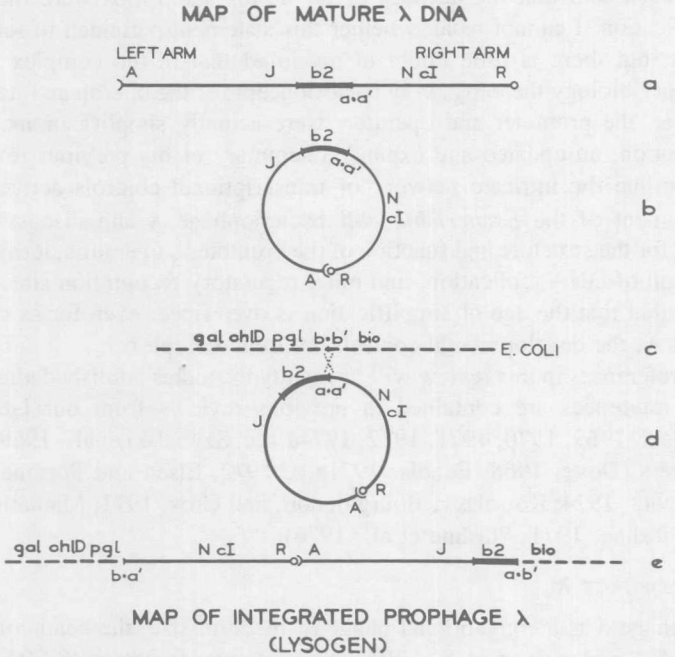


Fig. 1. Campbell's model for the integration (a→e) and excision (e→a) of the coliphage λ genome. Symbols *a.a'* and *b.b'* represent the attachment site on the viral (solid lines) and bacterial (broken lines) genomes, respectively (modified from Szybalski *et al.*, 1969).

lytic λ functions. The lysogenic response can occur if the repressor is produced early enough to block lytic development and hence to prevent death of the infected host cell.

In the lysogenic state the integrated phage genome, now denoted *prophage*, can be considered as a cluster of genes comprising about 1% of the host DNA. Since the λ prophage is an integral part of the bacterial genome, it is vertically inherited by all the bacterial progeny cells, a very efficient and harmless form of symbiotic propagation of the viral genome. Obviously, prophage propagation by this means would be terminated in the event of death of the lysogenic host cell. To ensure its survival, the λ prophage evolved a mechanism for sensing impending demise of the host and for entering the lytic cycle, which leads to production of a crop of *mature phage* particles. Conversion from the lysogenic (prophage) state into the lytic cycle is denoted *induction*, and its first step is the inactivation of the λ repressor protein.

Below I shall describe first the transcriptional controls in the repressed λ prophage and then the chain of events following induction of the lysogenic cell. A simplified genetic and transcriptional map of phage λ DNA is shown in figure 2.

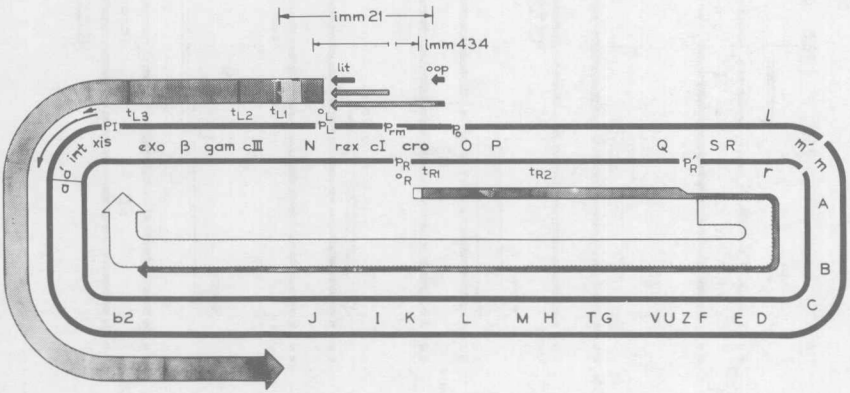
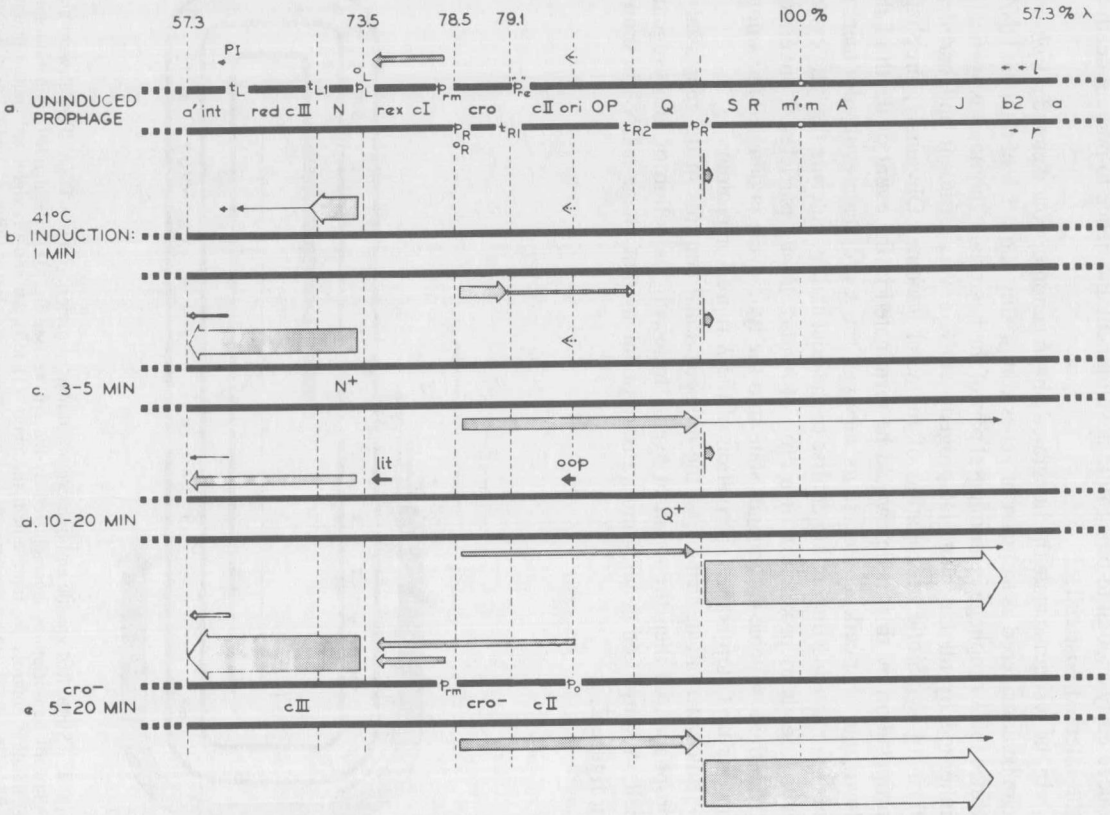


Fig. 2. Simplified genetic and transcriptional map of bacteriophage λ DNA. The leftward and rightward transcriptions are indicated by the arrows (see fig. 3). Five major termination-antitermination controls are represented: the major (1) leftward and (2) rightward transcriptions, initiated by the p_L and p_R promoters and prematurely terminated at the t_{L1} and t_{R1} terminators, respectively, are extended by the action of the *N* product; the (3) *oop* and the (4) *int* leader RNAs, initiated by the p_o and p_I promoters and prematurely terminated at the t_o and t_I terminators, respectively, are extended by the concerted action of the *c*II and *c*III products, to code for genes *c*I and *int*, respectively; (5) the 198-nucleotide RNA originated at the p_R promoter is elongated into the major late RNA by the action of the Q product. A detailed description of the various genes and recognition sites was compiled by Szybalski (1974b,c, 1976).


57.3
73.5
78.5
79.1
100 %
57.3 % λ
PI
a UNINDUCED PROPHAGE
a' int
red cIII
N
rex cI
cro
cII ori OP
Q
S R
m' m
A
J
b2
a
b INDUCTION: 41°C 1 MIN
c. 3-5 MIN
N+
lit
oop
d. 10-20 MIN
Q+
e 5-20 MIN cro−
cIII
cro−
cII

THE PROVIRUS STATE

To establish a provirus state (also denoted prophage or lysogenic state) two conditions must be met: (1) the expression of viral genes that code for products harmful to the host must be curtailed; and (2) the replication of the provirus genome must be carefully controlled and coordinated with that of the host. Under normal circumstances the λ prophage satisfies condition (1) by directing the synthesis of the repressor protein, the product of λ gene *c*I, which blocks directly the major leftward and rightward viral transcriptions and indirectly the autonomous λ DNA replication. Condition (2) is met by linear insertion of λ DNA into the host genome, which places the replication of the prophage entirely under the host DNA replication controls.

In other systems different means are known or could be imagined to satisfy conditions (1) and (2). For example, to meet condition (1) the repressor might be elaborated by the host, instead of by the virus; or the viral attachment site (symbol *a.a'*; fig. 1) might be located between the lytic viral genes and their promoter, with linear insertion of the provirus leading to "split" and thus inactivation of the viral operon. Condition (2) can often be satisfied by formation of an intricately controlled *plasmid,* rather than by integration. Even λ mutants that have lost the capacity to integrate (λN^-, λdv) can enter a fairly stable plasmid state, thanks to their capacity for autogenous regulation of the genes that control replication of the plasmid. This autogenous regulation is attributed to the product of gene *cro* (or *tof*), a second repressor of λ, whereas the repressor produced by gene *c*I is dispensable for maintenance of the plasmid state. The replication of such a plasmid cannot be either too rapid, since it would then outgrow and kill the host, or too slow, which would cause it to be progressively lost by dilution.

Fig. 3 (opposite). Schematic representation of the temporal sequence of transcriptional events in prophage λ. The leftward transcripts are coded by the *l* strand and drawn above λ DNA; the rightward transcripts are coded by the *r* strand and drawn below λ DNA. (a) Transcription in the uninduced prophage. The *c*I-*rex* transcripts correspond to 80–90% of the total prophage RNA. (b) Immediate-early transcription after induction. (c) Delayed-early transcription. (d) Late transcription. (e) Decontrolled transcription in an induced *cro*$^-$ mutant of λ. The prophage maps are not drawn to scale, with the immunity region expanded. The numbers in the top line indicate the positions of various sites in respect to the left end (0 %λ) and right end (100 %λ) of mature λ DNA (on the prophage map the 0 and 100 %λ termini are fused and represented by the 100 %λ point; see fig. 1a): *att* (57.3 %λ), p_I promoter (about 60.3 %λ; S. Hu, W. Szybalski and A. Campbell, unpublished data), t_{LI} (about 71.1 %λ; J. Salstrom and W. Szybalski, unpublished), s_L-$p_L o_L$ region (about 73.5 %λ), p_{rm}-$p_R o_R$ region (about 78.5 %λ), t_{RI} (about 79.1 %λ), p_o (about 80.2 %λ), *oop* RNA (79.9 to 80.1 %λ), *ori* site (about 81 %λ), t_{R2} site(s) (84 to 89 %λ), p_R' (or p_Q) (92 to 93 %λ). The width of the arrows is a measure of the rate of transcription. In the case of the 198 nucleotide 6 S RNA transcribed early (p_R'; figs. 3a–c), it was found by Dahlberg and Blattner (1973) that *in vitro* synthesis provides 10 to 20 times more of the 5′-proximal 15 nucleotide sequence (represented by the verical line in a-c) than of the total 6 S RNA (arrow under p_R'). Thus, p_R' is the strongest λ promoter but is immediately followed by strong termination signals that could apparently be overcome by the *Q* product, with resulting synthesis of late RNA (drawing d) (modified from Szybalski, 1974a).

TRANSCRIPTION IN THE PROPHAGE STATE

In the prophage state, in which the λ genome exists as an innocuous or even beneficial component of the bacterial chromosome, only one major operon of λ is transcribed. The host RNA polymerase recognizes the p_{rm} promoter and synthesizes the mRNA for genes *c*I and *rex*, copying the *l* strand of λ DNA (fig. 3a). The *c*I-*rex* mRNA is translated into the λ repressor protein (*c*I product) and the *rex* product. The repressor interacts with the o_L and o_R operators and blocks expression of all the major λ genes. Thus in the prophage state only about 4% (map position 74.3–78.4%λ; fig. 3a) of the prophage genome is transcribed. The λ repressor confers immunity against infection by λ phage, whereas the *rex* function blocks the development of certain mutants of the unrelated coliphages T1, T4 and T5. In this manner prophage λ pays tribute to the host by offering limited protection against several phages, a factor of possible evolutionary significance. The *rex* gene function might also be of some importance for efficient phage propagation or lysogenization under certain adverse conditions (Campbell and Rolfe, 1975), and it was even reported that a λ lysogen can have a selective advantage over nonlysogens by reproducing more rapidly (Edlin, Lin, and Kudrna, 1975). As described later, the *rex* product could also facilitate the natural induction.

The *c*I-*rex* transcription corresponds to about 90% of the total λ prophage transcription, with the remainder assigned to a few other sites, including the *oop* (traces of an about 78-nucleotide RNA) and *int* leader RNAs on the *l* strand and a 198-nucleotide RNA in the p_R' region on the *r* strand. However, these very minor transcriptions (also some in the b2 region; fig. 3a) appear to be of no physiological importance to the maintenance of the prophage state. As will be discussed later, the p_{rm}-promoted *c*I-*rex* transcription is controlled in both a positive and negative fashion.

INACTIVATION OF THE REPRESSOR PROTEIN

To induce phage development, it is necessary first to inactivate the λ repressor. To simplify this task, especially since *natural* (or *indirect*) induction is a slow and asynchronous process, many λ mutants have been isolated in which the *c*I protein is thermosensitive (ts). Heating these λ*c*Its lysogens to about 41°C inactivates the repressor but does not otherwise interfere with phage development. Since the active repressor has an oligomeric structure, its inactivation is probably associated with dissociation into *c*I subunits. This *direct* method of inactivation of the repressor is commonly practiced in the laboratory because it results in synchronous and almost instantaneous induction.

The *indirect* (or *natural*) method of prophage induction is based on the fact that transient interference with the host DNA synthesis by, e.g., irradiation, base analogs, or mitomycin treatment, results in a chain of events leading to the appearance of "special structures" (gaps; R. Sussman, personal communication) in the DNA that have acquired a low affinity for the ind^+ repressor of λ. The repressor is competitively scavenged from the o_L and o_R operators (Sussman and Ben Zeev, 1975), and, in addition, it is enzymatically cleaved (Roberts and Roberts, 1975). This indirect process requires over a half-hour and is inoperative if the bacteria carry mutations like *rec* A or *lex*, or if the phage has an ind^- mutation in gene *c*I. Apparently the *rec*A and *lex* products are involved in the creation of the "special structures" in the total DNA of the lysogen, and the ind^- mutation abolishes the repressor affinity for these structures. The indirect route of repressor inactivation is prevalent among many kinds of lysogenic strains as found in nature and, as already mentioned, is probably of evolutionary significance because it permits rescue of the prophage from a sickly or dying host cell by conversion into mature and infective phage particles.

Induction can also occur during conjugation when the prophage is transferred from the lysogen into a repressor-free receptor cell. This phenomenon is called *zygotic* induction.

TRANSCRIPTION AFTER INDUCTION OR INFECTION

The major events after λ prophage induction or after phage infection are (1) initiation of p_L and p_R promoted transcription (fig. 3b), (2) extension of this *immediate-early* transcription beyond the *t* terminators (*delayed-early transcription;* fig. 3c) and (3) appearance of a high level of p_R'-promoted *late* RNA (fig. 3d). Both the positive and negative controls of these events will be discussed. The immediate turnoff of the p_{rm}-*c*I-*rex* transcription after induction, the appearance of the new short *oop* and *lit* transcripts, and the turn-on of the *c*II-*c*III-dependent immunity transcription will be described next. The *c*II-*c*III-dependent *immunity-establishment* transcription (see the longest leftward p_0-*oop*-*c*II-*c*I-*rex* arrow in figure 4) is characteristic of the lysogenic response after infection, and is then replaced by the p_{rm}-promoted *immunity-maintenance* *c*I-*rex* transcription (fig. 3a). After induction, the immunity-establishment type of transcription is observed only in special (cro^-) mutants (fig. 3e).

As will be discussed later, this scheme is the first approximation of the actual events, which are really more complex as to be well in tune and quite responsive to the ever changing natural environment. However, we shall first

present an idealized picture of the chain of transcriptional events that follow λ prophage induction. Subsequently, we plan to discuss some of the events in more detail and describe the molecular mechanisms underlying the initiation, termination and antitermination processes.

IMMEDIATE-EARLY TRANSCRIPTION AFTER PROPHAGE INDUCTION

Inactivation of the λ repressor and its removal from the o_L and o_R operators permits the host RNA polymerase to initiate the leftward and rightward transcriptions at the p_L and p_R promoters (fig. 3b). The leftward transcription, however, does not proceed very far (about 2%λ), the bulk of it being terminated at the t_{L1} terminator. Similarly, most of the rightward transcription is blocked at the t_{R1} terminator, with only about an 0.5%λ length being transcribed, and the remainder of the t_{R1} readthrough is probably stopped at the t_{R2} terminator (fig. 3b). The host factor denoted *rho* (ρ) is instrumental in blocking transcription at the *t* sites. Immediate-early leftward transcription yields mRNA for gene *N*, and the bulk of the immediate-early rightward mRNA codes for the product of gene *cro* (or *tof*).

DELAYED-EARLY TRANSCRIPTION

The product of gene *N* acts as the antitermination factor. As will be outlined below, it interacts with the host RNA polymerase and abolishes the *rho*-imposed termination at the *t* sites. In this manner the leftward transcription extends from the p_L site to gene *int*, and the rightward transcription, which originates at p_R, covers genes *cro*, *O*, *P*. *Q*, and beyond (fig. 3c).

LATE TRANSCRIPTION

Although the p_R-initiated delayed-early transcription appears to extend beyond gene *Q*, it yields very little mRNA that codes for the head and tail genes *A* to *J*, barely enough to produce protein components for one phage particle per cell. To amplify the transcription in the *S-R-A-J* region, a special regulatory mechanism is provided. The product of gene *Q* permits the RNA polymerase to override the strong termination signals after the 15th and 198th nucleotide and to extend the p_R'-initiated minor rightward RNA, with a resulting massive rightward transcription of the *S-J* region (fig. 3d). Thus enough products are provided for about 100 or more phage particles per cell.

This orderly and sequential chain of transcriptional events results in expression of all the λ genes and should lead to production of a healthy crop of progeny phage. However, additional controls are required to make the process more efficient and more responsive to environmental factors. In the following

three sections I shall outline the interplay of genes *cro*, *c*I, and *N* in controlling the transcription promoted by p_L, p_R, and p_{rm}, the three modes of the immunity region transcription, and the role of transcription in the initiation of λ DNA replication.

REGULATION OF EARLY TRANSCRIPTION

The genome of λ codes for two repressor proteins, the products of genes *c*I and *cro*, and it is responsive to their action. As already mentioned, the *c*I product has three direct effects: it very efficiently blocks the transcriptions promoted by p_L and p_R, and has both a positive and a weak negative effect on the p_{rm}-*c*I-*rex* transcription. The product of gene *cro*, the "second repressor" or "antirepressor," also has three effects: it blocks quite effectively the p_{rm} and p_L-promoted transcriptions (Ai and Tof functions, respectively), and it depresses the p_R-promoted early rightward transcription (Tor function). These three functions of the gene *cro* product are shown schematically in figure 3. The Ai (=anti-immunity) function results in the disappearance of the *c*I-*rex* transcription (fig. 3a versus 3b-d), and the Tof and Tor functions depress the p_L-*N*-*int* and p_R-*cro*-*Q* transcription, respectively (thinner arrows in figure 3d than in figure 3c). Figure 3e schematically represents the decontrolled transcription in the induced *cro*$^-$ lysogen.

Since the p_L-promoted *in vivo* transcription is quite powerful and several of its products are toxic to the host and required only early in λ development, the Tof controlling mechanism is quite beneficial to an orderly replication and high λ yields. The Tor effect on the p_R-initiated rightward transcription is an example of an autorepression (autogenic control), since the *cro* protein is the product of the p_R-promoted operon and it regulates its own expression. This phenomenon might be quite important for λ DNA replication since genes *O* and *P*, which control λ DNA replication, are a part of the same operon. For instance, λ*N*$^-$*c*I$^-$ (or its p_R-*O*-*P* fragments, denoted λdv), can persist as autonomous plasmids and replicate in concert with the host. Autogenous control of the λ replication genes by the Tor function is probably the mechanism that (1) keeps λ replication in check during the lysogenic response, so as to allow effective lysogenization without killing the host cell, and which (2) in special cases permits establishment of a stable plasmid-carrier state by maintaining a precise and self-regulating balance between the replication of the λ plasmids and that of the carrier cells.

The negative regulation by the *cro* product is actually more complex than summarized above. It was shown recently that the *cro* product efficiently represses the leftward transcription of gene *N* only when RNA is synthesized by the *N*-modified RNA polymerase. (Hu, Salstrom, and Szybalski, 1975).