Enhancers and Eukaryotic Gene Expression

Current Communications
IN MOLECULAR BIOLOGY
Cold Spring Harbor Laboratory / 1983

Enhancers and Eukaryotic Gene Expression

Edited by

Yakov Gluzman
Cold Spring Harbor Laboratory

Thomas Shenk
State University of New York, Stony Brook

Titles in
Current Communications in Molecular Biology
PLANT INFECTIOUS AGENTS
ENHANCERS AND EUKARYOTIC GENE EXPRESSION

ENHANCERS AND EUKARYOTIC GENE EXPRESSION

International Standard Book Number 0-87969-161-1
Book design by Emily Harste
Printed in the United States of America

Front cover: Cloned fragments of MMTV DNA that contain specifically bound glucocorticoid receptor molecules (see Yamamoto et al., this volume). (Electron micrograph provided by B. Edgar and K.R. Yamamoto, University of California, San Francisco.)

Back cover: Phenotypic variation in the expression of SV40 and HSV-*tk* marker DNAs cotransfected into Rat-2 tk^- cells. Cotransformation of these two markers results in a low phenotypic expression of SV40 morphological transformation. Insertion of an LTR into the SV40 construction increases the cotransformation index almost to 1. (*Top row, left*) tk^+, morphologically transformed cells, $T\text{-}Ag^+$; (*right*) tk^+, morphologially nontransformed cells, $T\text{-}Ag^-$. (*Middle*) tk^+, split phenotype (morphologically transformed cells, $T\text{-}Ag^+$; morphologically nontransformed cells, $T\text{-}Ag^-$). (*Bottom row*) Sectored colony morphology; (*left*) tk^+, transformed and nontransformed cell sectors: dark sectors (morphologically transformed), $T\text{-}Ag^+$; light sectors (morphologically nontransformed), $T\text{-}Ag^-$; (*right*) 10× magnification of left panel (see Lusky et al., this volume). (Photographs provided by M. Kriegler and M. Botchan, Dept. of Molecular Biology, University of California, Berkeley.)

The individual summaries contained herein should not be treated as publications or listed in bibliographies. Information contained herein can be cited as personal communication contingent upon obtaining the consent of the author. The collected work may, however, be cited as a general source of information on this topic.

All Cold Spring Harbor Laboratory publications are available through booksellers or may be ordered directly from Cold Spring Harbor Laboratory, Box 100, Cold Spring Harbor, New York 11724.
SAN 203-6185

Conference Participants

Steven Beckendorf, Dept. of Molecular Biology, University of California, Berkeley

Arnold Berk, Molecular Biology Institute, University of California, Los Angeles

Paul M. Bingham, Dept. of Biochemistry, State University of New York, Stony Brook

Michael Botchan, Dept. of Molecular Biology, University of California, Berkeley

Kathryn Calame, Dept. of Biological Chemistry, Center for the Health Sciences, UCLA School of Medicine, Los Angeles, California

Mario R. Capecchi, Dept. of Biology, University of Utah, Salt Lake City

Pierre Chambon, Institut de Chimie Biologique, Faculté de Médecine de Strasbourg, France

William Dynan, Dept. of Biochemistry, University of California, Berkeley

Richard A. Fisher, Biogen, Inc., Cambridge, Massachusetts

Michael Fried, Imperial Cancer Research Fund Laboratories, London, England

Yakov Gluzman, Cold Spring Harbor Laboratory, New York

Chris Goff, Collaborative Research, Inc., Waltham, Massachusetts

Michael Green, Dept. of Biochemistry and Molecular Biology, Harvard University, Cambridge, Massachusetts

Peter Gruss, Institute of Microbiology, University of Heidelberg, Federal Republic of Germany

Leonard P. Guarente, Dept. of Biology, Massachusetts Institute of Technology, Cambridge

Gordon L. Hager, Laboratory of Tumor Virus Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland

Dean H. Hamer, Laboratory of Biochemistry, Cellular Regulation Section, National Cancer Institute, National Institutes of Health, Bethesda, Maryland

John A. Hassell, Dept. of Microbiology and Immunology, McGill University, Montreal, Quebec, Canada

Alan G. Hinnebusch, Dept. of Biology, Massachusetts Institute of Technology, Cambridge

Bruce H. Howard, Laboratory of Molecular Biology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland

Culai Hsiao, Central Research and Development Department, E.I. du Pont de Nemours & Company, Wilmington, Delaware

Thomas Kadesch, Dept. of Biochemistry, Stanford University School of Medicine, Stanford University Medical Center, California

Robert Kamen, Genetics Institute, Boston, Massachusetts

George Khoury, Laboratory of Molecular Virology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland

Arthur Levinson, Molecular Biology Department, Genentech, Inc., South San Francisco, California

Elwood Linney, La Jolla Cancer Research Foundation, Cancer Research Center, La Jolla, California
Frank McCormick, Cetus Corporation, Berkeley, California
Stella Mitrani-Rosenbaum, Laboratory of Pathology, Viral Oncology and Molecular Pathology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland
Joseph Nevins, Laboratory of Molecular Cell Biology, Rockefeller University, New York, New York
Alfred Nordheim, Dept. of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts
Vernon T. Oi, Dept. of Genetics, Stanford University School of Medicine, California
H.R.B. Pelham, MRC Laboratory of Molecular Biology, University Medical School, Cambridge, England
Alexander Rich, Dept. of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts
Bernard Roizman, Kovler Viral Oncology Laboratories, The University of Chicago, Illinois
Paolo Sassone-Corsi, Institut de Chimie Biologie, Faculté de Médicine de Strasbourg, France
Walter Schaffner, Institut für Molekularbiologie II der Universität Zürich, Switzerland
Thomas E. Shenk, Dept. of Microbiology, State University of New York, Stony Brook
David Tiemier, Molecular Biology Department, Monsanto Corporation, St. Louis, Missouri
Susumu Tonegawa, Center for Cancer Research, Massachusetts Institute of Technology, Cambridge, Massachusetts
Inder M. Verma, The Salk Institute, San Diego, California
T. Gordon Wood, Laboratory of Molecular Oncology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland
Keith R. Yamamoto, Dept. of Biochemistry and Biophysics, School of Medicine, University of California, San Francisco
Moshe Yaniv, Dept. of Molecular Biology, Institut Pasteur, Paris, France
Kai Zinn, Dept. of Biochemistry and Molecular Biology, Harvard University, Cambridge, Massachusetts

Preface

A conference on enhancers and eukaryotic gene expression was held at the Banbury Center of the Cold Spring Harbor Laboratory on April 3-6, 1983. Approximately 40 molecular biologists who have been studying the expression of a variety of eukaryotic genes presented and discussed the most recent advances in this field. The material presented at the conference consisted of two major topics: enhancers and regulated promoters.

Enhancer elements were originally discovered as novel elements contained in the SV40 early promoter. They have received considerable attention because of their ability to increase the transcriptional activity in *cis* of the bona fide SV40 promoter as well as of a variety of heterologous promoters. These elements have now also been found in the long terminal repeats of most retroviruses and polyoma, papilloma, BK, and Ad2 viruses as well as in immunoglobulin genes. Analysis of many inducible promoters has revealed that regulatory sequences are either present in the vicinity of promoters or overlapping them, and in a few cases it has been shown that these sequences act in a manner independent of orientation, a property that is characteristic of enhancer elements.

Enhancer and Eukaryotic Gene Expression summarizes the results of this conference in the form of extended abstracts provided by the speakers. These abstracts describe various experimental systems and include the majority of the known enhancers as well as a large number of inducible promoters from virus, yeast, *Drosophila*, mouse, and human origin. The book is intended to provide the interested reader with an overview of current work in this rapidly progressing field.

The support for this meeting was provided by Abbott Laboratories; Applied Molecular Genetics, Inc.; Biogen N.V.; Cetus Corporation; Collaborative Research, Inc; E.I. duPont de Nemours & Co.; Genentech, Inc; Lilly Research Laboratories; Molecular Genetics, Inc.; and Monsanto Company.

We wish to thank James D. Watson, Director of the Cold Spring Harbor Laboratory, for making available the facilities of the Banbury Center, as well as Dr. Mike Shodell and his staff of the Banbury Center and Nancy Ford, Doug Owen, Judy Cuddihy, and the Publications staff for their efforts in making possible a timely publication in a rapidly progressing area.

Yakov Gluzman
July 1983

Contents

Introduction

D.H. Hamer and G. Khoury

National Cancer Institute, National Institutes of Health
Bethesda, Maryland 20205

Transcription initiation is an early and critical event in eukaryotic gene expression. Changes in the rate and specificity of initiation play key roles in controlling cellular differentiation and development, adaptation to environmental conditions, and the fate of infecting viruses. What are the DNA sequences that determine how and when a gene is transcribed? What are the cellular factors that interact with these sequences to prompt initiation?

The development of eukaryotic gene transfer methods, together with in vitro transcription systems, has made an approach to the first question both feasible and popular. By constructing specifically mutated genes, then analyzing their expression in vivo and in vitro, it is possible to correlate the DNA sequence of a molecule with its transcriptional activity. Such studies have rapidly led to the discovery of a novel class of eukaryotic transcriptional control signals, called enhancers, and to insights into the regulated transcription of inducible genes. We will briefly review the basic methodologies used in these experiments and some of the early, published observations on enhancer elements and control sequences.

GENE TRANSFER AND IN VITRO TRANSCRIPTION

There are two basic procedures for studying the in vivo expression of cloned genes. In transient systems, the gene of interest is introduced into a population of cultured cells and its activity is assayed within a few hours to a few days. The original transient expression experiments utilized encapsidated SV40 recombinants that, at high multiplicity, could infect every cell in the culture. Subsequently, the procedure has been greatly simplified by the use of direct DNA transfections using plasmids grown in *Escherichia coli* (Mulligan and Berg 1980; Banerji et al. 1981; Mellon et al. 1981). Although only a small fraction of the cells take up and express the recombinant plasmid, transcription of the foreign gene can readily be detected by sensitive S1-nuclease or gel-transfer hybridization procedures, using highly radioactive DNA probes. Alternatively, if the control region of the gene under study is fused to the coding sequences of a gene such as herpes simplex virus thymidine kinase (HSV *tk*) or *E. coli* chloramphenicol acetyltransferase (*cat*), its activity can be monitored

by appropriate enzymatic assays. The obvious advantages of the transient assays are simplicity and rapidity. Furthermore, because the recombinant molecules remain episomal, many of the problems associated with random chromosomal integration are avoided. Disadvantages include a low efficiency of expression and the inability to conduct experiments over a period of more than a few days.

In long-term assays, one first isolates a clone of cells carrying the gene of interest in a stable fashion (Wigler et al. 1979). This usually involves the use of a second, selectable gene such as HSV *tk*, which is useful only in tk^- cell lines, or *E. coli* aminoglycoside phosphotransferase, which confers neomycin resistance in many different cell types. The selectable gene can be fused to the control region of the gene under study, present on the same plasmid as a distinct transcription unit, or cotransfected on a separate plasmid. In the latter case, investigators rely on the property of a small fraction of recipient cells to take up large quantities of DNA and on the high level of recombination between transfecting DNA molecules. Selected clones are grown in mass culture, checked for the presence of the nonselectable gene by DNA blotting, then evaluated for expression by the usual RNA and protein analyses. The major point in favor of long-term assays is that the cell lines can be grown indefinitely, thereby allowing one to study physiological phenomena and to isolate large quantities of the recombinant gene products. An important disadvantage is the tremendous variability observed among different clones carrying the same gene. This is generally attributed to differences in the "chromatin environment" of the foreign gene integration site. Although the molecular basis of this phenomenon remains unclear, it obviously makes quantitative interpretations of long-term experiments difficult.

The recent development of vectors based on bovine papilloma virus (BPV) offers some of the advantages of both the transient and long-term assay systems (Sarver et al. 1981). BPV can stably transform cultured mouse cells and is maintained as a nuclear episome. This insures that permanent cell lines are available for repeated experimentation and that the gene of interest is associated with a vector that may maintain a uniform chromatin structure accessible to experimentation. An additional advantage is the ease of recovering and cloning the episomal DNA for subsequent analyses.

The methods for introducing genes into yeast are considerably more advanced than those available in higher eukaryotes (for review, see Botstein and Davis 1982). Recombinant molecules can be propagated as plasmids at high or low copy number by using a variety of cloned DNA replication origins and centromeric se-

quences. Moreover, because yeast undergoes homologous recombination with relatively high efficiency, it is possible to direct the integration of a cloned gene to its natural site on the chromosome. This leads to a tandem duplication that, upon resolution by unequal crossing over, results in the exact replacement of the chromosomal gene with its cloned counterpart. It is clear that this is an important consideration in studying the expression of in-vitro-mutated genes and that the development of similar gene replacement methods for higher eukaryotes would be most desirable.

In vitro transcription provides some obvious and important alternatives to the in vivo gene transfer methods for studying gene regulation. The two major systems currently in use (Weil et al. 1979; Manley et al. 1980) consist of essentially whole cell lysates that may or may not require supplementation with purified RNA polymerase II, the enzyme responsible for transcribing all known eukaryotic coding genes. These systems have been used extensively to study the DNA sequence requirements for accurate in vitro initiation of transcription (for review, see Breathnach and Chambon 1981; Shenk 1981). Such studies have revealed two important regions: an AT-rich sequence approximately 30 nucleotides upstream from the initiation site (the "TATA" or "Goldberg-Hogness" box) and the sequences around the initiation site itself (the "cap" site). Mutations in these regions can alter both the specificity and efficiency of initiation. However, in most cases there seems to be little or no effect of sequences further upstream that are known to be critical in vivo. In the case of the sea urchin histone H2A gene (Grosschedl and Birnstiel 1982) and the adenoviral late genes (Hen et al. 1982), it has been possible to see an effect of upstream sequences by using supercoiled rather than linear templates, but the generality of this observation remains to be tested. Attempts to observe regulated in vitro transcription of inducible or repressible genes have been, for the most part, unsuccessful. A notable exception is the down-regulation of simian virus 40 (SV40) early transcription by the early viral protein, T antigen (Hansen et al. 1981; Myers et al. 1981). It is noteworthy that this effect is not seen in whole-cell lysates from early- versus late-infected cells but requires the addition of relatively large quantities of purified T antigen. Similar strategies may be required to observe the in vitro regulation of other repressible or inducible genes.

ENHANCER SEQUENCES

Enhancer sequences were discovered, and have been most thoroughly studied, in the small DNA tumor virus SV40. Figure 1

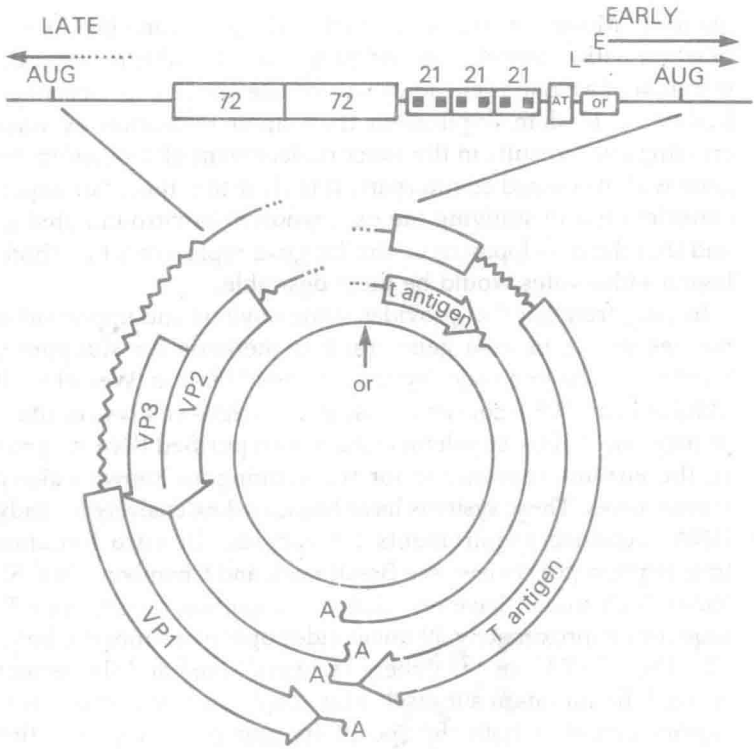

Figure 1 The transcription map and control region of SV40. The location of the origin of DNA replication (or) and the general structures of the predominant transcripts are diagramed. The early mRNAs encode the early proteins, small t antigen and large T antigen, whereas the late mRNAs encode the viral capsid proteins VP1, 2, and 3. The top part shows the viral control sequences lying between the early and late AUG initiation codons. This region contains a number of early transcriptional regulatory elements, including an AT-rich sequence (the "TATA" or "Goldberg-Hogness" box), three 21-bp repeats (which contain six GC-rich octanucleotides), and the two 72-bp repeat enhancer sequences. The early transcripts are initiated predominately at position E early in infection and shift to position L after DNA replication. The late viral transcripts have very heterogeneous 5′ ends. (The bottom part is redrawn from Tooze [1981], which is also the source for the BBB numbering system for SV40 nucleotides.)

shows a simplified map of the 5243-bp circular genome of this virus (for review, see Tooze 1981). There are two divergent transcription units: The early region is transcribed throughout the lytic cycle and in transformed cells, whereas the late region is transcribed only in permissive cells after the onset of viral DNA replication. The primary transcripts from both regions are differently spliced to yield

the two early mRNAs encoding the early proteins (T antigen and t antigen) and at least two late mRNAs encoding the late proteins (capsid proteins VP1, 2, and 3). The 400 bp between the early and late coding sequences contain the transcriptional control sequences of the virus (expanded portion of Fig. 1). Recent studies have focused on the role of this region in the initiation of early transcription, which occurs primarily at the position marked "Early E" (nucleotides 5232–5236) in the early stage of lytic infection and in transformed cells. Late in lytic infection the major cap site shifts approximately 40 bp upstream to position "Early L" (nucleotides 28–34).

The control region contains several important structural features, including two perfectly repeated 72-bp sequences between nucleotides 106 and 250, a series of three imperfect repeats of a 21-bp GC-rich sequence between positions 35 and 106, and an AT-rich sequence centered at position 20. Deletion of both 72-bp repeats leads to a drastic decrease in early region transcription as measured in transient assays (Benoist and Chambon 1981; Gruss et al. 1981). Surprisingly, transcription is largely restored by reinserting the 72-bp repeats at a variety of positions and in both possible orientations relative to the early region cap site (Moreau et al. 1981; Fromm and Berg 1982, 1983). The 72-bp repeat sequences also increase the transcription rate of many different cellular genes when they are present on the same plasmid (Banerji et al. 1981; Humphries et al. 1982; Treisman et al. 1982). These experiments laid the basis for our current working definition of an enhancer as a sequence that can potentiate the transcription of diverse genes relatively independent of position and orientation. The effects of the 72-bp repeats have been studied most extensively in transient expression assays using covalently closed, circular DNA molecules, but they also appear to be active in stably transformed cells (Capecchi 1980). However, no consistent effect of these sequences has yet been observed in in vitro transcription systems.

Only one of the two 72-bp repeat sequences of SV40 is required for full augmentation of early transcription in the usual assay systems (Subramanian and Shenk 1978; Gruss et al. 1981). Nevertheless, the fact that the enhancer regions of several other viruses also contain repeats suggests that sequence duplication may play an important role under certain conditions. The SV40 enhancer appears to function most efficiently as an intact unit, but fine-structure mapping indicates that the 3′ portion of the sequence (proximal to the early transcription unit) is less critical than the 5′ portion (Benoist and Chambon 1981; Fromm and Berg 1982; Weiher et al. 1983). The

discovery of limited sequence homologies between the enhancers of SV40 and other DNA and RNA tumor viruses has led to the proposal of a putative "consensus" enhancer sequence (Laimins et al. 1982; Weiher et al. 1983). Although this sequence by itself is unlikely to account for enhancement, it could serve as a "core" for interaction with common regulatory factors.

Mutations in the GC-rich and AT-rich sequences of the SV40 control region also affect the rate and, in some cases, specificity of initiation. These effects are observed both in vivo and in vitro (Mathis and Chambon 1981; Myers et al. 1981), but there is no evidence now that either of these regions can function in a position-independent fashion. Hence, these sequences appear more like "classical" promoter elements, of the sort extensively studied in bacteria. Similar promoter elements are present 5′ to many cellular genes as well.

Many laboratories are searching for, and discovering, enhancer sequences in animal viruses other than SV40; e.g., polyoma virus (de Villiers and Schaffner 1981; Tyndall et al. 1981), BPV (Lusky et al. 1983), adenovirus, and HSV. Perhaps the best-studied reside within the long terminal repeat sequences (LTRs) of the RNA tumor viruses or retroviruses (Laimins et al. 1982; Levinson et al. 1982; Jolly et al. 1983; Kriegler and Botchan 1983). Replacement of the 72-bp repeat sequences of SV40 with the corresponding 72/73-bp repeat sequences of Moloney murine sarcoma virus generates a viable, albeit crippled, SV40 virus (Levinson et al. 1982). It is also known that various LTRs, such as those of avian leukosis virus, can activate cellular proto-oncogenes following proviral integration (for review, see Temin 1982). The heterologous gene is often transcribed from its own initiation site, and activation occurs when the LTR is inserted either 5′ or 3′ to the proto-oncogene. Thus, the model previously described as "promoter insertion" might more appropriately be called "enhancer insertion."

The possibility that chromosomal genes possess their own enhancer elements, or reasonable facsimiles thereof, is also under intense investigation. The observation that expression of the sea urchin histone H2A gene is affected by a remote upstream sequence may have uncovered such a cellular enhancer (Grosschedl and Birnstiel 1980). Human genomic DNA sequences homologous to the SV40 72-bp repeats also behave like enhancer elements in some assay systems (Conrad and Botchan 1982). Another reason to suspect that some eukaryotic genes contain enhancers is the host-cell specificity manifested by certain viral enhancer elements. Thus, the SV40 enhancer is more efficient than the murine sarcoma virus and polyoma virus enhancers in monkey cells, whereas the converse is

true in mouse cells (de Villiers et al. 1982; Laimins et al. 1982). In this regard, it is particularly interesting that certain polyoma virus mutants capable of infecting undifferentiated mouse embryonal carcinoma cells contain DNA sequence alterations in what is now known to be the enhancer region of the virus (Katinka et al. 1980; Vasseur et al. 1980; Fujimura et al. 1981; Sekikawa and Levine 1981). Such experiments suggest the existence of species or cell-type-specific factors that interact directly with viral enhancers. It seems reasonable to speculate that such factors do not exist solely to accommodate invading viruses but that they also interact with chromosomal enhancer sequences.

CONTROL ELEMENTS OF INDUCIBLE GENES

Understanding how eukaryotic genes are turned on and off under different cellular conditions is a major challenge of modern molecular biology. Many of the early gene transfer experiments utilized genes that are normally expressed only in highly differentiated cell types (e.g., Mantei et al. 1979; Hamer et al. 1980). The observation that such genes are accurately and efficiently transcribed in heterologous, nondifferentiated tissue-culture cells indicates that the usual gene transfer methods are less than ideal for studying classical developmental regulation. On the other hand, there are many eukaryotic genes that, in response to environmental changes, must be rapidly activated or shut off without the accompaniment of cellular differentiation. Such inducible gene systems have proven more amenable to experimental analysis.

The heat-shock response is one of the best-studied examples of such an inducible eukaryotic gene system. Heat-shock genes have been found in all organisms examined, ranging from *E. coli* to man, and are transcriptionally activated in many cell types in response to elevated temperature or various other environmental stresses. A cloned *Drosophila* heat-shock gene (the hsp70 gene) retains its ability to be induced by heat when introduced into mammalian cells by cotransformation with a selectable marker (Corces et al. 1981; Burke and Ish-Horowicz 1982), transfection with an SV40 vector (Pelham 1982), or when microinjected into *Xenopus* oocytes (Bienz and Pelham 1982). Deletion mapping shows that the sequences 10–66 nucleotides upstream from the cap site are sufficient for heat-inducible transcription and that nucleotides between positions −47 and −66 are essential for the response (Pelham 1982). A synthetic version of the consensus region renders a nonrelated gene heat sensitive when placed at an appropriate distance from the cap site (Pelham and Bienz 1982). Thus, heat-shock gene transcription ap-