I. THE *E. COLI LAC* OPERATOR-REPRESSOR SYSTEM IS FUNCTIONAL FOR CONTROL OF GENE EXPRESSION IN ANIMAL CELLS.

II. MOLECULAR CLONING AND CHARACTERIZATION OF THE MOUSE SKELETAL MUSCLE (α) ACTIN GENE.

Thesis by

Mickey ChienTsung Hu

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1988

(Submitted May 11, 1988)

ACKNOWLEDGEMENTS

I wish to state my great gratitude to my advisor, Norman Davidson, not only for his guidance in research, including support, understanding, criticism, and scientific writing, but particularly for his always warm concern for my well-being. Also, I would like to thank my professors who offered support (scientific or nonscientific), suggestions, and/or equipment: Judy Campbell, Peter Dervan, Henry Lester, Jean-Paul Revel, Jack Richards, and Barbara Wold.

Many friends have helped me to make this thesis possible in either scientific or personal matters. I am grateful to all members, past and present, of ND's laboratory as well as some people in the laboratories at Caltech. Specifically, I thank Marie Krempin for her assistance with cell culture. Terry Snutch for his instruction on *Xenopus* oocyte microinjection methods and for providing lots of healthy oocytes. Frank Preugschat for his instruction on oocyte nuclear injections and Sandy Sharp for her help in my first year in ND's laboratory and the excellent collaboration on the actin gene project at the early stage. Nevis Fregien for his instruction on the Ca-phosphate DNA transfection technique and valuable input in the actin gene project. Al Goldin for his instruction and help on his version of IBM PC, Frank Calzone for his instruction on ^{35}S-DNA sequencing while the commercial sequencing kit was not available, and for the updated DNA-protein binding assay. Also, thanks to K. Blackburn for her tissue-culture hood and coffee supply, etc. To all those whom I have not thanked explicitly, blame not my memory, but the lack of time and space. Thanks also to the National Institutes of Health and Amgen, Inc. for research funding and pocket

money.

In particular, I wish to state my gratitude to Lei Yu for his exceptional help after I suffered from an automobile accident earlier this year. His help was more than what a good friend would offer. (I need not describe it here since I will never forget it.) In this particular event, I also thank Marie for her moral and temporary financial support; K. and ND for their moral support and for making one or more phone calls for me, Maribelle and Sabrina for their help at home and for receiving a few phone calls for me.

I am always grateful to my parents, my family, and Maribelle. Finally, I would like to thank my wife, Sabrina Chen, who always listened very patiently to my complaints when I was disappointed, since life at Caltech was sometimes full of rude awakenings. Without her faith and love, none of this would have been possible.

ABSTRACT

We have investigated the use of the *E. coli lac* operator-repressor system to regulate the expression of genes introduced into mammalian cells by gene transfer. We find that the bacterial *lac* repressor protein encoded in a suitable expression vector is synthesized in mammalian cells in culture, assembles into a tetramer, enters the nucleus to some extent, and represses expression of another gene that has one or several *lac* operator sequences inserted into any one of several sites in the promoter region of the gene. Derepression can be achieved by exposure of the cells to IPTG. From a practical point of view of an inducible genetic switch, this system confers an induction level of somewhere between 10- and 20-fold in most of the cases we have tested. That is not better than those that have been achieved with heat shock, mouse mammary tumor virus, and metallothionein promoters. There may, however, be some situations and some promoters for which the use of the *lac* operator system is advantageous. We have also shown that this *lac* control system can be used to regulate the expression of genes introduced into *Xenopus* oocytes by micro-injection.

At present, we have been trying to revise this system by using a newly developed promoter containing a symmetric *lac* operator sequence inserted at the various strategic points within the human metallothionein II_A promoter and enhancer regions, which consist of several positive control elements, in order to achieve induction ratios of a factor of 100 or more. We have also generated a new *lac*I gene which encodes a repressor containing at its carboxyl terminus the nuclear localization signal of the SV40 large-T antigen. Overall, by combining

the newly developed *lac* control promoters with the new repressor producing cell lines, we hope to generate an inducible expression system with large induction ratios that can be used as a general genetic switch in the future.

In section II, the nucleotide sequence of the mouse skeletal muscle (α) actin gene is presented and discussed.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION TO SECTION I

The *E. coli lac* operator-repressor system is functional for control of gene expression in animal cells.

The specific aim of the first (and major) part of this thesis is to develop a new inducible eukaryotic expression system that can be used to achieve high levels of expression of gene products of interest, by gene transfer in animal cells and with a high induction ratio. There are two basic kinds of applications of this system that we foresee: (a) the direct usefulness for the controlled expression of gene products in animal cells, and (b) for the study of one aspect of the mechanism of enhancer function in eukaryotic cells.

Reason (a) can be subdivided into two parts. First, for genetic engineering purposes, one often wishes to produce useful proteins, such as rare hormones or potent growth factors, by expression in animal systems. In some such cases, the gene product is toxic to the cells. An appropriate strategy then is to place the gene under the control of an inducible promoter, grow the cells up to large numbers while the gene is inactive, and then induce expression. Second, in basic science one would like to be able to study the mechanism of action of any endogenous gene product, especially regulatory molecules (including proto-oncogenes, kinases, growth factors, etc.) in a biological system of interest, by turning the gene on and off at will. For example, if one can block the expression, in a controlled manner, of a potent endogenous oncogene by inducible antisense RNA expression, one has the ideal control between plus and minus expression for observing effects of the oncogene.

Explanation of the second motivation (b) requires some discussion of enhancers and of our present knowledge of their mechanism of function. One of the central problems to understand in eukaryotic

molecular biology is the mechanisms by which specific genes are expressed in a tissue-specific manner or are actived in response to extracellular inducers. Essentially, two DNA sequence elements are required for the regulation of gene transcription in eukaryotes: promoters (including proximal and distal elements), which are required for accurate and efficient initiation of transcription from genes, and enhancers, which are needed to increase the rate of transcription from promoters. Promoters are located immediately upstream from the start site of transcription, and act in a position-dependent fashion (for reviews, see Dynan and Tjian, 1985; McKnight and Tjian, 1986; Jones et al., 1988), whereas enhancers can act on *cis*-linked promoters at great distances (a few kilobases away) in an orientation-independent manner and can also function downstream from the transcription unit (for reviews, see Gluzman, 1985; Serfling et al., 1985; Ptashne, 1986). The available data are consistent with the current view that promoters and enhancers operate by *trans*-acting protein factors that specifically interact with these *cis*-acting sequence elements and stimulate transcription *in vivo* and *in vitro* (for reviews, see McKnight and Tjian, 1986; Maniatis et al., 1987; Jones et al., 1988).

In recent years, inducible promoters and enhancers have been identified for a number of genes including genes for heat shock (Wu, 1984; Topol et al., 1985; Bienz and Pelham, 1986), metallothionein (Mayo et al., 1982; Karin et al., 1984a, 1984b; Friedman and Stark, 1985; Haslinger and Karin, 1985; Serfling et al., 1985; Imbra and Karin, 1987), interferon-α (Ryals et al., 1985) and -β (Fujita et al.,

1985; Goodbourn et al., 1985), and c-*fos* (Treisman, 1985). In addition, steroid-inducible enhancers have been identified in the long terminal repeat sequences of the mouse mammary tumor virus (see Ringold, 1983, for review; Chandler et al., 1983; Ponta et al., 1985) and the Moloney murine sarcoma virus (DeFranco and Yamamoto, 1986; Miksicek et al., 1986), and in association with many cellular genes (Karin et al., 1984; Renkawitz et al., 1984; Slater et al., 1985; Klein-Hitpass et al., 1986; Seiler-Tuyns et al., 1986; Jantzen et al., 1987). Inducible enhancer elements are generally capable of activating heterologous promoters. In fact, usually only very short DNA sequences within enhancers are necessary for activation of transcription from nearby promoters. For instance, a 12-base pair DNA sequence element from the human metallo-thionein gene is sufficient to confer metal inducibility on a heterologous promoter (Searle et al., 1985), and a 14-base pair DNA sequence element from the *Drosophila* hsp70 gene is sufficient to confer heat inducibility on a heterologous promoter in mammalian cells (Pelham and Bienz, 1982).

Numerous DNA binding and footprinting experiments have provided strong evidence that inducible enhancer elements specifically interact with cellular protein factors. For example, first, the best understood *trans*-acting regulatory factors in higher eukaryotes are the steroid receptor proteins that bind to the steroid-inducible enhancers and promote transcription (see Yamamoto et al., 1985, for review). It has been shown that cotransfection of plasmids expressing receptor proteins and a plasmid containing a steroid-inducible enhancer into receptor-negative cell lines led to the activation of transcription (Miesfield

et al., 1986; Druege et al., 1986; Giguere et al., 1986), suggesting that binding of steroid receptor-hormone complex to the enhancer is required for steroid-dependent gene regulation. Second, a protein factor that is capable of binding heat-shock response elements (HSEs) has been identified in yeast (Wiederrecht et al., 1987), *Drosophila* (Parker and Topol, 1984; Wu, 1984), and man (Kingston et al., 1987). The HSE-binding factor has been highly purified and shown to stimulate transcription from the hsp70 promoter *in vitro* (Parker and Topol, 1984) and therefore has been designated the heat-shock transcription factor (HSTF). However, the mechanism by which *trans*-acting protein factors bind to enhancer elements and act on the transcriptional machinery, including RNA polymerase II, to activate transcription is not understood for any inducible enhancer.

Let us assume that a specific protein binds to a distant enhancer element and then interacts with a protein complex, including RNA polymerase itself, at the promoter. Then there are two basic kinds of mechanisms by which the protein at the enhancer site can act at a distance on the promoter complex: (a) action through space via "looping", (b) action through the chain via "scanning". Ptashne (1986) has summarized the arguments for the looping model. This model has the advantage of intuitive plausibility. However, Courey et al. (1986) have offered evidence that enhancer function can be blocked by inserting a short segment of DNA containing intercalated photo-crosslinked psoralen molecules between enhancer and promoter, suggesting a through the chain mechanism (scanning model).

Our idea is to place a *lac* operator sequence between the enhancer

and the promoter elements. If action is through the chain, then the presence of *lac* repressor binding to the operator should block the action of the enhancer. Addition of IPTG should reestablish communication between the enhancer and promoter elements. On the other hand, it is unlikely that a *lac* repressor bound to an operator would prevent interaction via a looping mechanism if all the elements are suitably spaced. Our specific plans are to investigate this problem by inserting the *lac* operator between the glucocorticoid responsive element (GRE, an inducible enhancer) and the promoter region of the human metallothionein II_A gene (see below).

Let us return to a more detailed discussion of the use of the *lac* operator-repressor system for control of eukaryotic gene expression after gene transfer. There are of course other already available eukaryotic inducible promoter/enhancer systems (described above). These have some disadvantages that make it worthwhile to test new approaches. For example, the metal-ion-inducible metallothionein promoter and the glucocorticoid-inducible mouse mammary tumor virus promoter suffer often from the existence of a relatively high level of constitutive expression, and with the addition of inducing agents, they suffer from a limited inducibility or are restricted to a relatively narrow range of host cells. Also, the inducing agents, which include heavy metals, steroid hormones, interferons, and heat treatment, may interfere with normal metabolism and cell proliferation. And, although the heat-shock promoter has been suggested to be a general method for inducibly expressing proteins in mammalian cells (Dreano et al., 1986; Wurm et al., 1986), it has been reported recently that heat shock transiently

accelerates the degradation of the long-lived fraction of cellular proteins in some mammalian cells (Parag et al., 1987).

The specific aim of the first section of this thesis is thus to establish a new eukaryotic expression system that can be optimally applied to the large-scale expression of gene products of interest in mammalian cells. The approach of this project is to use the *E. coli lac* repressor and the *lac* operator to control expression of genes introduced into mammalian cells by gene transfer. The general background of this approach is described by the following. In the *E. coli lac* operon, genes encoding enzymes essential for lactose metabolism are under the negative control of a tetramer repressor molecule consisting of four identical polypeptide chains, each of molecular weight 38 kilodalton. The *lac* repressor can bind specfically to a cognate *lac* operator sequence downstream of the *lac* promoter. When bound to the operator, the repressor can completely inhibit transcription from the *lac* promoter in *E. coli*. Allolactose and its analogues, such as isopropyl β-D-thiogalactoside (IPTG), can bind to repressor protein and dramatically decrease the affinity of repressor to the operator so that they activate transcription of the *lac* promoter. Overall, the *lac* operator-repressor system is one of the most thoroughly studied and best understood examples of a protein-nucleic acid interaction that regulates transcrip-tion of a gene in prokaryotic systems (Beckwith and Zipser, 1970; Miller and Reznikoff, 1980).

We show in Chapter 2 that the bacterial *lac* repressor protein encoded in a suitable expression vector is synthesized in mammalian cells in culture, assembles into a tetramer, enters the nucleus to some

extent, and represses expression of another gene that has one or several *lac* operator sequences inserted into any one of several sites in the promoter region of this plasmid. Derepression can be achieved by exposure of the cells to IPTG. Although a significant level of derepression (60%-80%) can be achieved with IPTG, it is not fully reversible. From a practical point of view of an inducible genetic switch, this system confers an induction ratio somewhere between 10- and 20-fold in most of the cases we have tested. That is not better than those that have been achieved with heat shock, mouse mammary tumor virus, and metallothionein promoters. There may, however, be some situations and some promoters for which the use of the *lac* operator system is advantageous. At present, we have been trying to revise this system by combining it with other positive control elements so that we may get dynamic control over a factor of 100 or more (see below).

Additionally, we have investigated the use of this *lac* control system to regulate the expression of genes introduced by microinjection into *Xenopus laevis* oocytes. A primary goal of this study was to demonstrate that this system could be used as a genetic switch in frog oocytes, just as in *E. coli* and in mammalian cells (Hu and Davidson, 1987; Brown et al., 1987). The ability to control the transcription of a gene in oocytes by using the *lac* repressor and hybrid promoters containing the *lac* operator should be useful for a wide variety of purposes. Moreover, one potentially rewarding feature of this system is that it may provide a method the experimenter can use to control the expression of genes that play a role in early embryonic development. Ultimately, it may provide a method for the control of tissue-specific

gene expression in transgenic animals.

As described in Chapter 3, with a single *lac* operator sequence inserted in a control region of the test plasmid injected into *Xenopus* oocytes, a high level of repression is achieved with the *lac* repressor. Repression can be achieved by premixing repressor with plasmid prior to nuclear injection, or by injection of repressor into the oocyte cytoplasm before injection of the plasmid into the oocyte nucleus. A low level of induction can be achieved by addition of the inducer IPTG to the external medium; a much higher level of induction is obtained by injection of IPTG into the oocyte cytoplasm. Furthermore, the repressor protein does not affect expression of a test plasmid that does not contain the operator sequence.

It is clear that there is room for revision and improvement within this *lac* control system. First, we have been trying to introduce other positive control elements, including the metal responsive element (MRE) and the glucocorticoid responsive element (GRE), into this control system to achieve higher induction ratios upon the addition of the extracellular inducers (e.g. heavy metals + IPTG or dexamethasone + IPTG). The rationale of this approach is that by using a combination of a positive and negative control system, we likely will get multiplicative effects upon the addition of inducers because the positive activation should be able to occur independently and concurrently with the relief of repression from the negative control. The human metallothionein II_A (hMT-II_A) gene provides a useful model for this purpose. This gene is actively expressed in most cell types, and its transcription is regulated in response to a variety of environmental stimuli,

which includes heavy metal ions, steroid and polypeptide hormones, growth factors, cytokines, and phorbol ester tumor promoters (Karin et al., 1984a, 1984b; Friedman and Stark, 1985; Imbra and Karin, 1987). These effects are mediated by specific regulatory proteins that interact with distinct *cis*-acting elements in the 5'-flanking region of the gene (Karin et al., 1984b; Lee et al., 1987a, 1987b; Angel et al., 1987). In addition to one GRE and four MREs, the 5'-flanking region contains four other elements involved in controlling the basal level of expression: a TATA box, a GC box, and two basal level enhancer elements (BLEs) (Karin et al., 1987). Recently, we have successfully constructed a number of new expression vectors which consist of the hMT-II_A promoter with a single symmetric *lac* operator sequence inserted at the various strategic points within the promoter and enhancer regions, as shown in Figure 1 (page 16). It should be noted that a perfectly symmetric *lac* operator segment (Sadler et al., 1983; Simons et al., 1984) has been chosen for use here because it has been shown that it binds the *lac* repressor about 8-10 times more tightly than does the natural *lac* operator sequence, which has been used in our previous studies. Hopefully, further work may reveal that this new *lac* control system can achieve induction ratios of a factor of 100 or more.

Second, as mentioned previously, since the majority (≈90%) of the *lac* repressor synthesized in the mammalian cell line LI-1 which we have characterized is located in the cytoplasm, it is desirable to generate a new repressor that will localize primarily in the nucleus. In eukaryotes, most regulatory proteins (DNA-binding proteins) are transported from their cytoplasmic site of synthesis and are localized

primarily in the nucleus. This relatively strict nuclear accumulation is attributed to the presence in the protein of a distinctive stretch of amino acids that acts as a nuclear localization signal (see Dingwall, 1985, for review). Accordingly, as another venture of revising this *lac* control system, we have generated a new *lac*I gene that encodes a repressor containing at its carboxyl terminus the nuclear localization signal of the SV40 large-T antigen (Kalderon et al., 1984). As depicted in Figure 2 (page 17), a stretch of DNA sequence encoding the short amino acid sequence of Pro-Lys-Lys-Lys-Arg-Lys-Val has been engineered on the 3' end of the wild type *lac*I gene by using oligonucleotide-directed mutagenesis (Zoller and Smith, 1984). This gene has been stably transfected into three mammalian cell lines, including mouse Ltk$^-$, rat XC, and Rat-2(tk$^-$) cells. Experiments are currently under way to isolate and characterize the cell lines that synthesize high levels of repressor protein. Moreover, the distribution between cytoplasm and nucleus of the engineered *lac* repressor containing the nuclear localization signal will be determined by the immuno-blot (Western-blot) analysis. Overall, by combining the newly developed *lac* control systems with the new repressor producing cell lines, we hope to generate an inducible expression system with large induction ratios that can be used as a general genetic switch in the future.

REFERENCES

1. Angel, P., Imagawa, M., Chiu, R., Stein, B., Imbra, R.J., Rahmsdorf, H.J., Jonat, C., Herrlich, P., and Karin, M. (1987) *Cell* 49, 729-739.

2. Beckwith, J.R., and Zipser, D. (1970) *The Lactose Operon*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

3. Bienz, M., and Pelham, H.R.B. (1986) *Cell* 45, 753-760.

4. Brown, M., Figge, J., Hansen, U., Wright, C., Jeang, K., Khoury, G., Livingston, D.M., and Roberts, T.M. (1987) *Cell* 49, 603-612.

5. Chandler, V.L., Maler, B.A., Yamamoto, K.R. (1983) *Cell* 33, 489-499.

6. Courey, A.J., Plon S.E., and Wang, J.C. (1986) *Cell* 45, 567-574.

7. DeFranco, D., and Yamamoto, K.R. (1986) *Mol. Cell. Biol.* 6, 993-1001.

8. Dingwall, C. (1985) *Trends Biochem. Sci.* 10, 64-66.

9. Dreano, M., Brochot, J., Myers, A., Cheng-Meyer, C., Rungger, D., Voellmy, R., and Bromley, P. (1986) *Gene* 49, 1-8.

10. Druege, P.M., Klein-Hitpass, L., Green, S., Stack, G., Chambon, P., and Ryffel, U. (1986) *Nucleic Acids Res.* 14, 9329-9337.

11. Friedman, R., and Stark, G. (1985) *Nature* 314, 637-639.

12. Fujita, T., Ohno, S., Yasumitsu, H., and Taniguchi, T. (1985) *Cell* 41, 489-496.

13. Giguere, V., Hollenberg, S.M., Rosenfeld, M.G., and Evans, R.M. (1986) *Cell* 46, 645-652.

14. Gluzman, Y. (ed.) (1985) *Eukaryotic transcription: the role of cis- and trans-acting elements in initiation*. Cold Spring Harbor

Laboratory, Cold Spring Harbor, N.Y.

15. Goodbourn, S., Zinn, K., and Maniatis, T. (1985) *Cell* 41, 509-520.
16. Haslinger, A., and Karin, M. (1985) *Proc. Natl. Acad. Sci. USA* 82, 8572-8576.
17. Hu, M.C-T., and Davidson, N. (1987) *Cell* 48, 555-566.
18. Imbra, R.J., and Karin, M. (1987) *Mol. Cell. Biol.* 7, 1358-1363.
19. Jantzen, H.-M., Strahle, U., Gloss, B., Stewart, F., Schimid, W., Boshart, M., Miksicek, R., and Schutz, G. (1987) *Cell* 49, 29-38.
20. Jones, N.C., Rigby, P.W.J., and Ziff, E.B. (1988) *Genes Devel.* 2, 267-281.
21. Kalderon, D., Roberts, B.L., Richardson, W.D., and Smith, A.E. (1984) *Cell* 39, 499-509.
22. Karin, M., Haslinger, A., Holtgreve, H., Cathala, G., Slater, E., and Baxter, J.D. (1984a) *Cell* 36, 371-379.
23. Karin, M., Haslinger, A., Holtgreve, H., Richards, R.I., Krauter, P., Westphal, H.M., and Beato, M. (1984b) *Nature* 308, 513-519.
24. Karin, M., Haslinger, A., Heguy, A., Dietlin, T., and Cooke, T. (1987) *Mol. Cell. Biol.* 7, 606-613.
25. Kingston, R.E., Schuetz, T.J., Larin, Z. (1987) *Mol. Cell. Biol.* 7, 1530-1534.
26. Klein-Hitpass, L., Schorpp, M., Wagner, U., and Ryffel, G.U. (1986) *Cell* 46, 1053-1061.
27. Lee, W., Haslinger, A., Karin, M., and Tjian, R. (1987a) *Nature* 325, 368-372.
28. Lee, W., Mitchell, P., and Tjian, R. (1987b) *Cell* 49, 741-752.
29. Maniatis, T., Goodbourn, S., and Fischer, J.A. (1987) *Science* 236,

1237-1245.

30. Mayo, K.E., Warren, R., and Palmiter, R.D. (1982) *Cell* 29, 99-108.

31. McKnight, S., and Tjian, R. (1986) *Cell* 46, 795-805.

32. Miesfield, R., Rusconi, S., Godowski, P.J., Maler, B.A., Okret, S., Wikstrom, A.-C., Gustafsson, J.-A., and Yamamoto, K.R. (1986) *Cell* 46, 389-399.

33. Miksicek, R., Heber, A., Schmid, W., Danesch, U., Posseckert, G., Beato, M., and Schutz, G. (1986) *Cell* 46, 283-290.

34. Miller, J.H., and Reznikoff, W.A. (1980) *The Operon*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

35. Parag, H.A., Raboy, B., and Kulka, R.G. (1987) *EMBO J.* 6, 55-61.

36. Parker, C.S., and Topol, J. (1984) *Cell* 37, 273-283.

37. Pelham, H.R.B., and Bienz, M. (1982) *EMBO J.* 1, 1473-1477.

38. Ponta, H., Kennedy, N., Skroch, P., Hynes, N.E., and Groner, B. (1985) *Proc. Natl. Acad. Sci. USA* 82, 1020-1024.

39. Ptashne, M. (1986) *Nature* 322, 697-701.

40. Renkawitz, R., Schutz, G., von der Ahe, D., and Beato, M. (1984) *Cell* 37, 503-510.

41. Ringold, G.M. (1983) *Curr. Topics Microbiol. Immunol.* 106, 79-103.

42. Ryals, J., Dierks, P., Ragg, H., and Weissman, C. (1985) *Cell* 41, 497-507.

43. Sadler, J.R., Sasmor, H., and Betz, J.L. (1983) *Proc. Natl. Acad. Sci. USA* 80, 6785-6789.

44. Searle, P.F., Stuart, G.W., and Palmiter, R.D. (1985) *Mol. Cell. Biol.* 5, 1480-1489.

45. Seiler-Tuyns, A., Walker, P., Martinez, E., Merillat, A.-M., Givel,

F., and Wahli, W. (1986) *Nucleic Acids Res.* 14, 8755-8770.

46. Serfling, E., Jasin, M., and Schaffner, W. (1985) *Trends Genet.* 1, 224-230.

47. Simons, A., Tils, D., von Wilcken-Bergmann, B., and Muller-Hill, B. (1984) *Proc. Natl. Acad. Sci. USA* 81, 1624-1628.

48. Slater, E.P., Rabenau, O., Karin, M., Baxter, J.D., and Beato, M. (1985) *Mol. Cell. Biol.* 5, 2984-2992.

49. Topol, J., Ruden, D.M., and Parker, C.S. (1985) *Cell* 42, 527-537.

50. Treisman, R. (1985) *Cell* 42, 889-902.

51. Wiederrecht, G., Shuey, D.J., Kibbe, W.A., and Parker, C.S. (1987) *Cell* 48, 507-515.

52. Wu, C. (1984) *Nature* 309, 229-234.

53. Wurm, F.M., Gwinn, K.A., and Kingston, R.E. (1986) *Proc. Natl. Acad. Sci. USA* 83, 5414-5418.

54. Yamamoto, K.R. (1985) *Annu. Rev. Genet.* 19, 209-252.

55. Zoller, M., and Smith, M. (1984) *DNA* 3, 479-488.

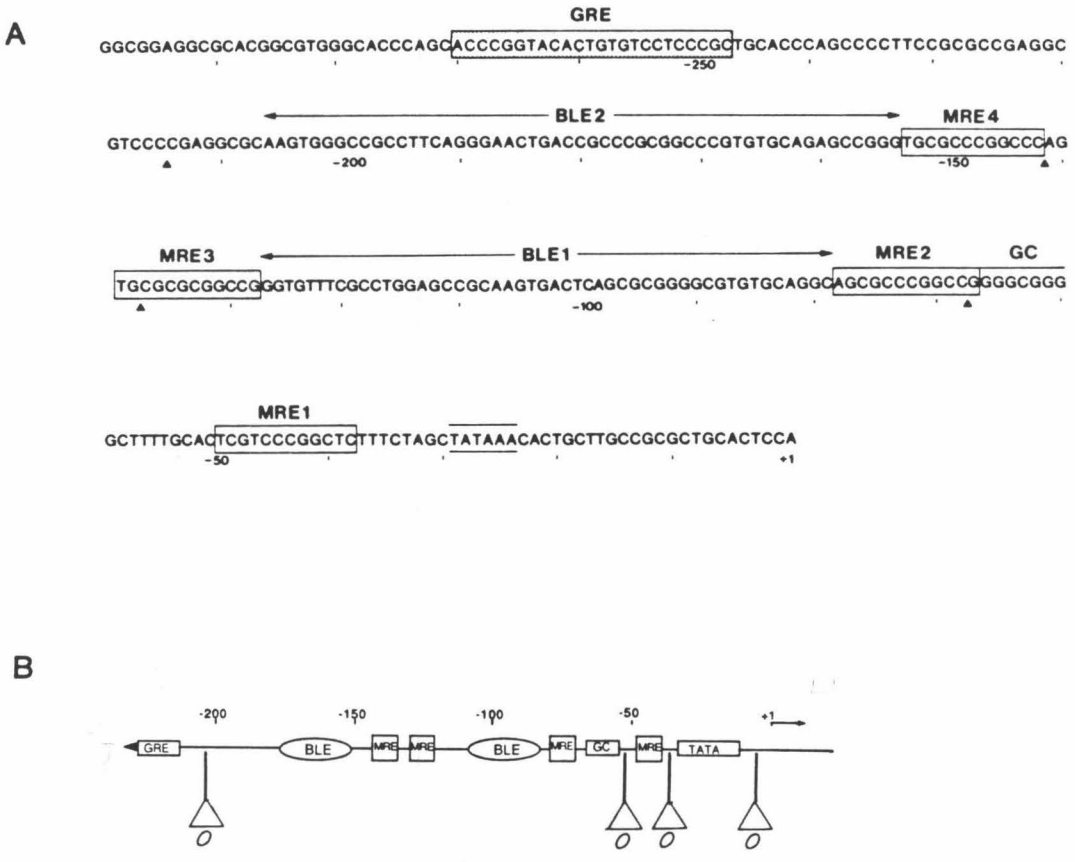


Figure 1. Structure of the hMT-II_A control region. (A) Sequence of the hMT-II_A 5′ control region from position +1 (the major start site of transcription) to position -300. (B) Schematic diagram of the *lac* operator insertions in the promoter region. BLE, basal level enhancer; GC, GC-box element; *O*, the *lac* operator.

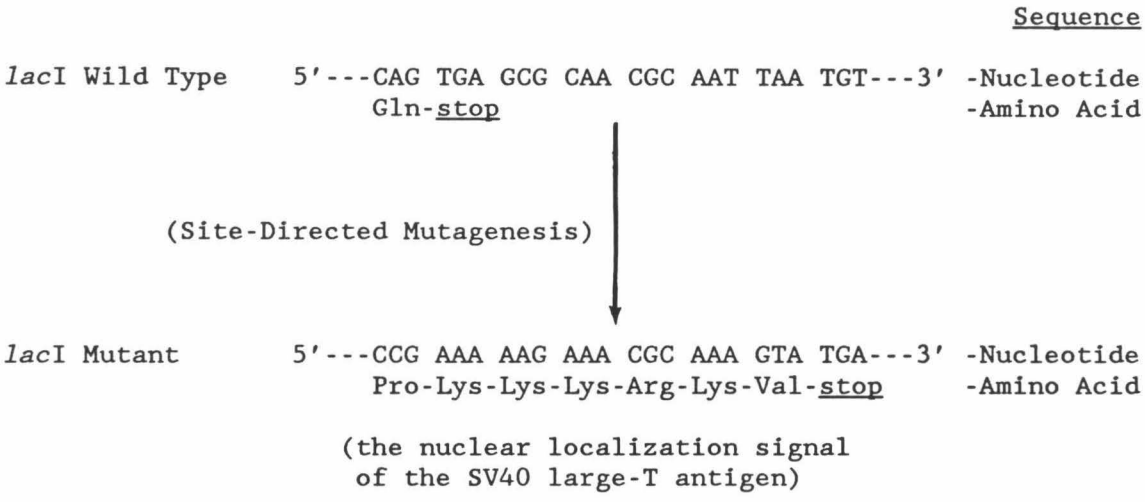


Figure 2. Oligonucleotide-directed mutagenesis of the *lac*I gene. The mutant *lac*I gene encodes a repressor containing at its carboxyl terminus the nuclear localization signal of the SV40 large-T antigen.

CHAPTER 2

The inducible *lac* operator-repressor system is functional in mammalian cells.

Mickey C-T. Hu and Norman Davidson

Cell 48:555-566 (1987)

The Inducible *lac* Operator–Repressor System Is Functional in Mammalian Cells

Mickey C.-T. Hu and Norman Davidson
Division of Chemistry
California Institute of Technology
Pasadena, California 91125

Summary

We have investigated the use of the Escherichia coli *lac* operator-repressor system to regulate expression of transfected genes in mammalian cells. We show that *lac* repressor produced in mouse L cells by transfection of a *lacI* expression vector blocks transcription of an MSV-CAT fusion gene when the *lac* operator is inserted at any one of the following sites within the promoter region: between the initiation codon (ATG) and the transcription start site; between the transcription start and TATA box regions; or upstream of the TATA box region. This last result suggests that the repressor may prevent protein-protein interactions involved in transcription activation. The inducer IPTG causes a marked derepression of CAT expression. The *lac* repressor-operator complex may be useful as an on/off "switch" in the regulation of gene expression for gene transfer experiments.

Introduction

Emerging evidence indicates that one of the major mechanisms for control of gene expression in eukaryotes, just as in prokaryotes, involves binding of a regulatory protein to a specific DNA sequence that it recognizes (for reviews see Dynan and Tjian, 1985; McKnight and Tjian, 1986). Generally speaking, these regulatory mechanisms are somewhat more complex in eukaryotes than in prokaryotes in that they more frequently involve interactions between several different proteins each binding to a distinct short sequence in the regulatory region of the gene (Ptashne, 1986; Davidson et al., 1986).

Repression by binding of the *lac* repressor to the *lac* operator and derepression by action of inducers such as isopropyl β-D-thiogalactoside (IPTG) is one of the most thoroughly studied and best understood examples of a protein–nucleic acid interaction that regulates transcription of a gene in Escherichia coli (Beckwith and Zipser, 1970; Miller and Reznikoff, 1980). The goal of the present research was to determine whether a suitably engineered *lac* repressor–operator system would be functional in animal cells just as in E. coli. If effective, such a system would be a useful addition to the several eukaryotic inducible promoter systems now used for achieving regulated expression of genes introduced into eukaryotic cells by gene transfer. These latter include heat shock inducible promoters (Wu, 1984; Topol et al., 1985), the glucocorticoid-inducible mouse mammary tumor virus long terminal repeat (LTR) (see Ringold, 1983, for review), the metal-ion-inducible metallothionein promoter (Mayo et al., 1982), and poly(IC)-inducible interferon promoters (Goodbourn et al., 1985; Ryals et al., 1985). In addition, information about steric interactions between two eukaryotic regulatory proteins, each binding to its specific sequence in a promoter region, might be provided by observing the effects of *lac* repressor binding to an operator sequence inserted in the vicinity of the two interacting regulating sequences.

The *lac* repressor is a homotetramer containing four polypeptide chains, each of molecular weight 38 kd. To be functional in a eukaryotic cell, this protein would have to self-assemble after synthesis of the monomer units in the cytoplasm, migrate to the nucleus, and bind to an operator sequence that was present in some sort of a chromatin structure characteristic of eukaryotic DNA. To be useful the system would have to be inducible by IPTG added to the external medium. It was not known beforehand whether all this would occur. To our knowledge, the only instance in which a prokaryotic regulatory protein has been shown to bind to its specific recognition sequence in a eukaryotic cell is for the E. coli *lexA* protein to bind to its operator in yeast (Brent and Ptashne, 1984).

Results

Overview

Our overall strategy is depicted in Figure 1. By mutation of the initiator GTG to ATG, the *lacI* coding sequence has been modified so that it is expected to be translated in a eukaryotic cell. It was then inserted into a eukaryotic expression vector driven by the Rous sarcoma virus LTR (RSV LTR) as a promoter. A mouse L cell line denoted LI-1, in which functional repressor is expressed at moderate levels, was selected (Figure 1A).

The *lac* operator sequence has been inserted into selected sites around the promoter element or transcription start site of a standard DNA construction, pSM12, in which the chloramphenicol acetyltransferase (CAT) coding sequence is transcribed under the control of a chimeric promoter consisting of part of the SV40 early promoter coupled to the Moloney sarcoma virus (MSV) enhancer (Laimins et al., 1984). SV40 RNA processing signals are provided at the 3′ end of the gene. This particular construction was chosen because the effect of the MSV enhancer in augmenting expression had already been studied. Furthermore, a series of plasmids with the enhancer in different positions relative to the promoter has been made and studied by Laimins et al. (1984); we anticipate, in future studies, measuring the effect of the position of the *lac* operator relative to the enhancer and the promoter in these several constructions.

As summarized in Figure 1B, we have studied CAT gene expression, both by enzyme assays and RNA blot analysis, after introducing the plasmids by a transient transfection method into the repressor-positive LI-1 cells and, as a control, into the parental L cells. The effect of the inducer IPTG has been measured. Several specific CAT construc-

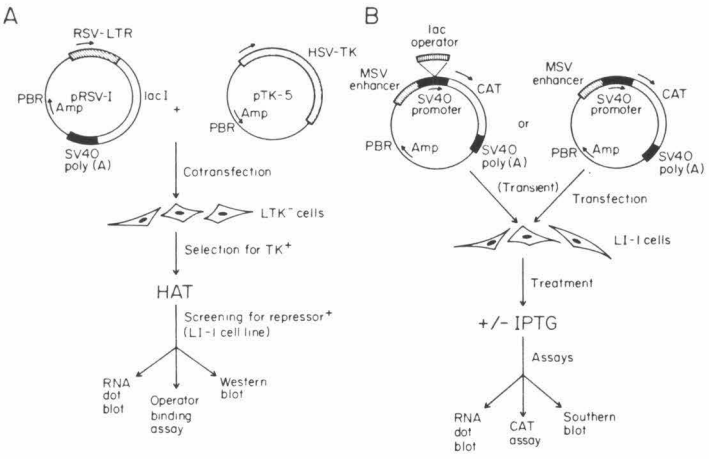


Figure 1. Schematic Representation of the Expression of *lac* Repressor and Its Functional Assay

(A) An expression vector containing an altered *lacI* gene under the control of the RSV LTR was cotransfected into LTK⁻ cells with a plasmid pTK-5, which carries the HSV-1 *tk* gene. The transfected cells were cultured in HAT selective medium, and clonal isolates were screened for the presence of stably expressed *lacI* mRNA and repressor protein (see Experimental Procedures). (B) In the functional assay, an MSV–CAT fusion gene or its derivatives containing *lac* operator sequence was transfected into the repressor producing cells (LI-1). The cells were subsequently cultured in the presence or absence of IPTG. The transient expression of the CAT gene was monitored by the CAT assay and RNA blot analysis, and nuclear DNA uptake was quantitated by Southern blot analysis.

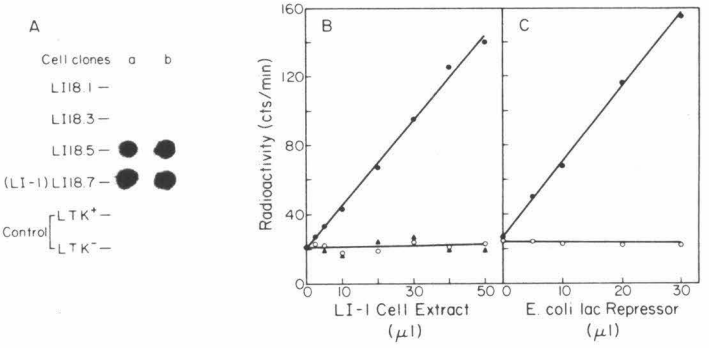


Figure 2. Expression of *lacI* mRNA and *lac* Repressor Protein

(A) RNA dot-blot analyses were performed using total cytoplasmic RNA extracted from L cells (HAT-resistant clones) that had been transfected 2 weeks (a) and 2 months (b) previously with the plasmids pRSV-I and pTK-5. RNA extracted from nontransfected LTK⁻ cells and those transfected with pTK-5 plasmid (TK⁺ cells) were also performed as controls. Equal amounts (4 μg per dot) of each RNA sample were transferred to a filter and were subsequently hybridized with an SP6-generated antisense *lacI* probe. (B) Aliquots of crude cytoplasmic extracts (approximately 2.4 mg/ml) from LI-1 cells and ^{32}P-end-labeled 40 bp operator DNA (2×10^{-12} M) were equilibrated in the presence of 20 μg/ml sonicated calf thymus DNA in standard binding buffer at room temperature (Lin and Riggs, 1972). The reaction was divided into three portions, and each sample was filtered through a nitrocellulose membrane filter (BA85, Schleicher & Schuell); the filters were washed with buffer and counted for radioactivity. No background radioactivity has been subtracted. (●) No IPTG; (○) in the presence of 5 mM IPTG; (▲) cellular extracts from LTK⁻ cells used as a control in the absence of IPTG. (C) Standard binding curve for E. coli *lac* repressor (approximately 27 ng/ml) and operator DNA. The binding assay was performed as described above. No background radioactivity has been subtracted. (●) No IPTG; (○) in the presence of 5 mM IPTG.

tions have been studied. As noted above, the parental plasmid, pSM12, contains a gene consisting of an MSV enhancer, the SV40 promoter, the CAT coding region, the SV40 small t intron, and an SV40 poly(A) addition site. This construction is expected to be transcriptionally active in repressor-positive cells because it does not carry the *lac* operator sequence. The constructions with an operator insert contain: one or several tandem operator insertions between the initiation codon and the transcription start point, designated as pSMAO1, pSMAO2, etc.; one or several tandem operator inserts between the transcription start point and the TATA box region, designated as pSMBO1, pSMBO2, etc.; or one or two inserts between the TATA box region and upstream elements (SV40 21 bp repeats), designated as pSMCO1 and pSMCO2, respectively.

It seemed plausible on general grounds that two or more tandem operator insertions might have a greater effect on gene expression than would a single sequence. This expectation is strengthened by the report that the half-lives of complexes of repressor bound to operator-containing plasmids increase with increasing numbers of tandem operators per plasmid (Sadler et al., 1980).

Functional *lac* Repressor Can Be Expressed in Mouse L Cells

The initiation codon of the E. coli *lacI* gene is GTG (Steege, 1977) and is not expected to function in eukaryotes. (It is of interest to note that changing the GTG initiation codon of a *lacZ* fusion gene into ATG actually results in a higher level of expression in E. coli in vivo and in vitro systems [Looman and van Knippenberg, 1986].) Therefore, in order to achieve a moderately high level of expression of the *lac* repressor in animal cells, we have converted the initiation codon of the *lacI* gene to ATG by site directed mutagenesis, and inserted the coding region into a eukaryotic expression vector driven by the RSV LTR and followed by SV40 splicing and polyadenylation signals, to give the plasmid pRSV-I (see Experimental Procedures).

Mouse LTK$^-$ cells were cotransfected with a plasmid (pTK-5) encoding the herpes simplex virus-I thymidine kinase gene (*tk*) plus a 50-fold excess of pRSV-I by the standard calcium phosphate coprecipitation procedure (Wigler et al., 1979; see Experimental Procedures). Stable transfectants were selected by growth in HAT-medium (see Experimental Procedures; Szybalska and Szybalski, 1962). Clonal isolates were screened for the presence of stably expressed *lacI* mRNA and repressor protein. Many clones expressed the repressor gene by these criteria. RNA dot-blot data for several clones and appropriate negative controls are shown in Figure 2A. Further studies have been confined to the cell line LI-1, which is the highest level expresser of the clones isolated.

The tetramer repressor protein produced in E. coli is capable of binding to the operator sequence in vitro. To determine whether repressor that is functional by this criterion is produced in the transfected animal cells, we performed a filter-binding assay for repressor–operator DNA complexes (Lin and Riggs, 1972). Crude extracts prepared from the cotransformed clones were incubated with the purified ^{32}P-end-labeled operator fragment (40 bp), in the presence or absence of IPTG, prior to filtering through a nitrocellulose filter. In addition, purified *lac* repressor protein was used to standardize the operator-binding curve. Figure 2B shows a filter-binding assay of cell extracts from the *lacI*$^+$ LI-1 cells, and demonstrates that these cells synthesize functional repressor protein that specifically binds the operator fragment in the presence of a vast excess of calf thymus DNA just as does the purified E. coli protein. By comparison of the linear binding curves of the extracts and of the purified protein, we estimate that there are 2.5 ($\pm$0.5) $\times$ 10^4 functional repressor molecules per cell. The molecular weight of the tetramer is 150 kd (Gilbert and Müller-Hill, 1970); therefore, assuming 10^{-9} gram of protein per cell, the functional repressor constitutes ca. 6 $\times$ 10^{-6} by weight of total cell protein.

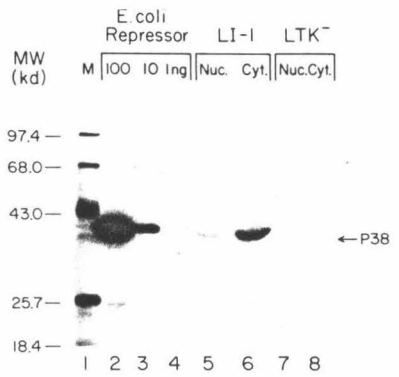


Figure 3. Quantitative Western Blot Analysis of *lac* Repressor in LI-1 Cells

Crude cytoplasmic and nuclear extracts from LI-1 and LTK$^-$ cells (2 $\times$ 10^6) were electrophoresed through an SDS–polyacrylamide gel and were transferred to nitrocellulose. The blot was probed with anti-repressor monoclonal antibody B-2 (Sams et al., 1985) followed by ^{125}I-labeled sheep anti-mouse immunoglobulin. After autoradiography, the bands were quantitated by densitometric scanning. The migration of *lac* repressor monomer (P38) is indicated. The positions of ^{14}C-labeled molecular weight standards are shown in the left margin. Cyt., cytoplasmic fraction; Nuc., nuclear fraction; M, marker proteins.

The *lac* Repressor Is Distributed Between Cytoplasm and Nucleus

Although the *lac* repressor is a well-characterized DNA-binding and regulatory protein in E. coli, it was not known whether it could penetrate the nuclear envelope and how it would distribute between cytoplasm and nucleus in animal cells. To address this question, Western blot analyses were performed on extracts of separated cytoplasmic and nuclear fractions of LI-1 cells, using a repressor-specific monoclonal antibody B-2 (Sams et al., 1985; see Experimental Procedures).

The immunoblots in Figure 3 show that a repressor polypeptide of approximately 38 kd comigrating on an SDS gel with the monomer derived from purified E. coli repressor is present in LI-1 cells but not in LTK$^-$ cells. About 10% of the total repressor was found in the nuclear fraction and 90% in the cytoplasm. The estimates from these gels of total repressor polypeptide present agree well with the amount found by operator binding; thus, most of the repressor polypeptide in LI-1 cells is present as functional tetramer. We therefore estimate that about 2500 repressor molecules are present in the nucleus, where they might be able to act as negative regulatory molecules.

The measured ratio of nuclear to cytoplasmic repressor could be quantitatively distorted by nuclear leakage or disruption and/or by adherence of cytoplasmic material to the external nuclear surface. The gene expression experiments reported below, as well as experiments with stable transformants and experiments by Xenopus oocyte injection methods (unpublished results), all show that there is some biologically active repressor in the nucleus.

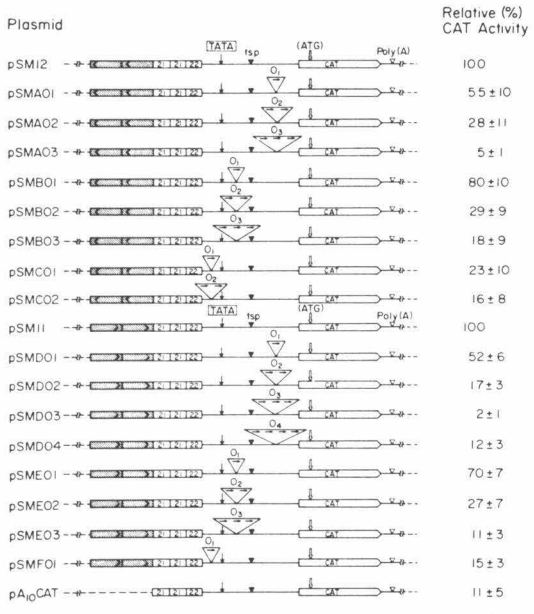


Figure 4. Structural and Functional Properties of Operator Insertions in the Promoter Regions

The organization of the SV40 early promoter region in parental plasmids (pSM12 and pSM11) and in their derivatives containing operator insertions is depicted. The hatched boxes to the left of the SV40 promoter illustrate the MSV enhancer (the 73/72 bp tandem repeats; Laimins et al., 1984). The SV40 21 bp repeats, TATA box, transcription start point (tsp), initiation codon (ATG), CAT-coding sequences (boxed region), and poly(A) addition site are indicated. The positions of single and multiple tandem operators (denoted as horizontal arrows) are shown as triangular inserts, and the copy number of the operator insert is indicated at the top of each triangle: O_1 (one operator), O_2 (two tandem operators), O_3 (three tandem operators), O_4 (four tandem operators). Individual cultures of LTK$^-$ cells were transfected with 4 μg (at 2 μg/ml) of the indicated plasmid DNA by using the DEAE-dextran method, and cell extracts were prepared for assaying CAT activity at 48 hr after transfection. CAT assays were performed at least three times, using equal amount of protein from the cell extracts, under linear assay conditions. The relative CAT activity is quantified as percentage of parental plasmid pSM12 CAT activity. The standard deviations shown on the right were calculated using data from four independent transfection experiments. Transfections and assays were repeated by using two different preparations of the plasmids that showed less than 50% CAT activity in transient expression in LTK$^-$ cells. The control CAT expression was also performed using $pA_{10}CAT$, the plasmid that contains the SV40 promoter elements but no enhancer sequences. The finite but low level of expression observed from this control plasmid has been observed by others (N. Fregien, personal communication). In our hands, the expression ratio for $pA_{10}CAT$ relative to pSM12 varies with amount of DNA and with host cell type. We suspect that transcription from $pA_{10}CAT$ may be due to a cryptic promoter within the vector, as we observe for pUCOCAT (constructed by N. Fregien), which has no eukaryotic promoter (see also Lopata et al., 1986; and Fregien and Davidson, 1986).

Effects of Operator Insertions on Gene Expression in the Absence of Repressor

Before studying the possible effects of repressor binding to operator sequences in the several plasmid constructions in vivo, it was necessary to determine the extent to which operator inserts affect gene expression in the absence of repressor. Accordingly, transient expression studies were carried out with these plasmids in LTK$^-$ cells. Figure 4 shows that insertion of a single operator between the initiation codon and the transcription start (pSMAO1 and pSMDO1) caused a 45%–48% decrease in CAT activity with respect to the parental plasmid (pSM12 and pSM11, respectively); two tandem operators (pSMAO2 and pSMDO2) caused a 72%–83% decrease in CAT activity; and three tandem operators (pSMAO3 and pSMDO3) almost eliminated CAT activity. Similarly, insertion of a single operator between the transcription start and the TATA box (pSMBO1 and pSMEO1) decreased CAT activity approximately 20%–30%; two tandem operators (pSMBO2 and pSMEO2) decreased CAT activity about 70%; and three tandem operators (pSMBO3 and pSMEO3) decreased CAT activity about 80%–90%. In contrast, insertion of a single operator between the TATA box and the upstream SV40 21 bp repeats (pSMEO1 and pSMFO1) resulted in a 77%–85% reduction in CAT activity. These latter results are in agreement with a previous study showing that insertion of DNA segments of increasing length between 21 bp repeats and the TATA box region leads to a significant decrease in the amount of RNA transcript initiated at the major early-early start sites (Takahashi et al., 1986). (It may be noted, in Figure 4, that we have studied operator insertions into constructions pSM11 and pSM12 of Laimins et al. (1984) with both orientations of the enhancer relative to the transcription direction. Our later studies in LI-1 cells have been restricted to a subset of these plasmids.)

The *lac* Repressor-Operator Complex Functions As a Regulatable Switch in Animal Cells

To examine the functional capability of the expressed repressor as a negative transcription regulator in mammalian cells, we performed CAT assays with extracts of LI-1 cells obtained after transfection with the reporter genes (see Experimental Procedures). As shown in Figure 5A, derivatives of the parental plasmid pSM12 containing a single operator between the initiation codon and the transcription start (pSMAO1, lane 2), between the transcription start and the TATA box regions (pSMBO1, lane 4), and between the TATA box and the SV40 21 bp repeats (pSMCO1, lane 6) were repressed 8–12-fold in the repressor-positive LI-1 cells relative to the same plasmids in the repressor-negative L cells. All pSM12 promoter derivatives containing two tandem operators (Figure 5A, pSMAO2 and pSMBO2) were repressed more efficiently (about 30-fold) than corresponding derivatives that contained only a single operator. In summary, in the presence of repressor, CAT enzymatic activity was reduced by 87%–92% for the single-operator-inserted genes and by 96%–98% for the two-operator-inserted genes.

To determine whether CAT gene expression was regulated at the level of transcription, we isolated total cytoplasmic RNA from transiently transfected LI-1 cells and determined the concentrations of CAT RNA by RNA dot-blot hybridization (Figure 6B). The relative amounts of CAT-specific RNA paralleled the levels of CAT enzyme activity in the transfected cells.

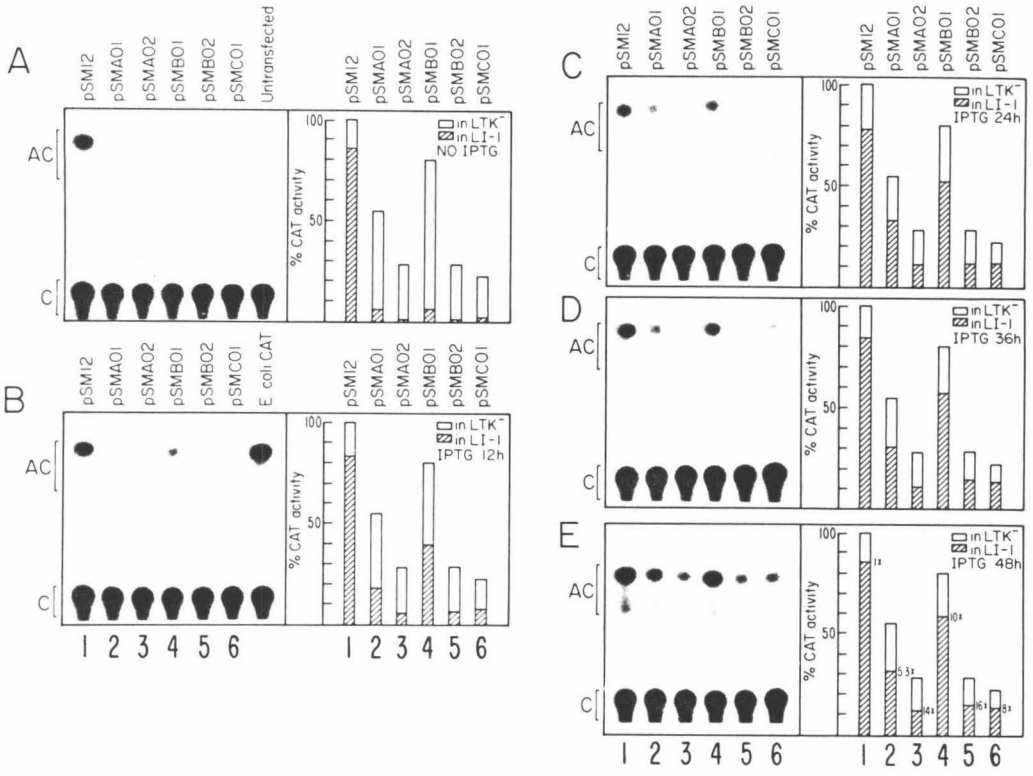


Figure 5. Repression and Induction of CAT Activity in Repressor-Producing Cells (LI-1)

Subconfluent LI-1 cells were transfected with 4 μg (at 2 μg/ml) of plasmid DNA, as indicated in each lane, by using the DEAE-dextran protocol. After transfection, cells were cultured for 48 hr in galatose-substituted DME media (see Results and Experimental Procedures) under the following conditions: (A) no IPTG was included; (B) IPTG was added to a 20 mM final concentration at 36 hr after transfection, and cells were continuously cultured for 12 hr; (C) IPTG added at 24 hr after transfection, and cells were cultured for another 24 hrs; (D) IPTG added at 12 hr after transfection, and cells were cultured for additional 36 hr; and (E) IPTG added right after transfection, and cells were cultured for 48 hr. Analyses of CAT activity were carried out 48 hr following transfection as previously described. The relative CAT activity is normalized as a percentage of the parental plasmid's (pSM12) CAT activity in LTK⁻ cells. Shown to the right of these CAT assays (A–E), the level of CAT expression from the indicated plasmid constructs in LI-1 cells (hatched bars) was compared with that in LTK⁻ (empty bars) which is shown in Figure 4. The magnitude of induction in 48 hr as shown in (E) was individually quantified by dividing the % CAT activity of each in (E) by that observed in (A) in LI-1 cells. As an additional negative control, when the operator was inserted downstream of the poly(A) site within the vector, CAT expression was not repressed in LI-1 cells.

These experiments are most plausibly interpreted as showing that binding of the repressor to an operator sequence inserted upstream of the transcription start point at the several sites that we have used inhibits either binding of RNA polymerase II or successful initiation of transcription. Inhibition by repressor bound to the operator at a site that begins 42 nucleotides downstream of the start point may be due to steric interference with initiation or, more plausibly, may interfere with propagation of the transcription complex. The latter view is consistent with the evidence that the *lac* repressor–operator complex is an efficient terminator of transcription both in vivo and in vitro in a prokaryotic expression system (Deuschle et al., 1986).

In E. coli the β-galactoside IPTG is an efficient inducer of *lac* gene expression from the repressed state. Since we have shown that repressor protein can act on operator sequences in animal cells, we next asked whether IPTG in the external medium can relieve this repression. We therefore measured CAT enzyme activity as a function of time after the addition of IPTG to the medium. Transfected LI-1 cells were grown in the galactose-substituted DME medium (see below) in the presence of 20 mM IPTG for 12 to 48 hr. At various times thereafter, cytoplasmic extracts were prepared, and CAT enzyme activity was measured. As shown in Figures 5B–5E, treatment of the transfected cells with IPTG induces expression of CAT activity. Induction occurred rather slowly, being half-maximal after 12 hr (Figure 5B) and nearing plateau by 24 hr (Figure 5C). The level of induction is ca. 5-fold for the construct pSMAO1 (Figure 5E, lane 2), 14-fold for pSMAO2 (lane 3), 10-fold for pSMBO1 (lane 4), 16-fold for pSMBO2 (lane 5), and 8-fold for pSMCO1 (lane 6). However, induction levels never returned to the fully derepressed level estimated from expression in LTK⁻ cells. These experiments were carried out with galactose rather than glucose in the tissue culture medium, because galactose is a weak inducer of *lac* repressor whereas glucose is a weak anti-inducer in bacteria and in vitro (Barkley and Bourgeois, 1980; Riggs et al., 1970). The magnitude of induction in the presence of a given concentration of IPTG in the galactose-substituted

DME media was about 2-fold higher than that in the glucose-containing DME media. There was no detectable induction in the galactose-substituted DME media in the absence of IPTG. In addition, we note that IPTG had no effect on the growth of LI-1 or LTK$^-$ cells.

The effects of IPTG were also monitored by measuring CAT mRNA levels. The RNA dot-blot hybridizations (Figure 6C) show that the amounts of CAT-specific RNA induced corresponded well with the levels of the CAT enzyme induced in transfected cells. The induction ratios in cells transfected with plasmids containing two tandem operators (pSMAO2 and pSMBO2) are greater than those for cells transfected with plasmids containing a single operator (pSMAO1 and pSMBO1). As previously noted, the repressed levels of expression are always lower for the plasmids containing two inserted operators than for the ones with a single operator. In neither case, however, does the maximally induced state reach the level observed for the same plasmids in LTK$^-$ cells, where there is no repressor.

In summary, these results show that the several MSV–CAT fusion genes carrying *lac* operators at different positions within the promoter regions were repressed by the *lac* repressor expressed in LI-1 cells in the absence of IPTG and were derepressed by induction with IPTG.

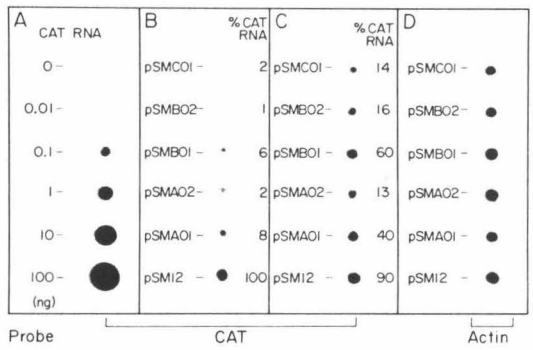


Figure 6. Analysis of CAT RNA Synthesized in the Absence or Presence of IPTG in LI-1 Cells

RNA dot-blot analyses were performed using RNA extracted from LI-1 cells that had been transfected 48 hr previously with the indicated plasmid constructs. RNA samples (4 μg per dot) were transferred to filters and were subsequently hybridized with an SP6-generated antisense CAT probe that hybridizes only to CAT-coding sequences or with an SP6-generated antisense actin probe (Hu et al., 1986). (A) Nontransfected LTK$^-$ cell RNA (2 μg) plus 0, 0.01, 0.1, 1, 10, or 100 ng of CAT RNA that was made by transcription in vitro of an SP6 clone (p64-CAT, which gives sense-strand RNA). (B) RNA from transfected LI-1 cells cultured in the absence of IPTG. (C) From a different transfection, RNA from LI-1 cells cultured in the presence of IPTG as described for Figure 5E. (D) A control experiment of (B) that was probed with an SP6-generated antisense actin probe. Controls show that the amount of total cellular RNA used per dot contributes negligibly to the signal in this assay. The relative CAT RNA in the respective transfectants was calculated after scintillation counting of each sample and comparing the radioactivities to that of the parental plasmid's CAT RNA.

Operator Inserts Immediately Downstream of the TATA Box Shift the Transcription Start Site

The current view on the function of the TATA element in higher eukaryotes is that it acts as a selector for transcriptional initiation sites by defining a fixed number of base pairs downstream at which RNA polymerase begins transcription. Consistent with this view, deletions of the TATA element result in heterogeneous initiation, and deletions that remove sequences downstream from a TATA element result in mRNA initiated from a new site (Grosschedl and Birnsteil, 1980; Benoist and Chambon, 1981; Mathis and Chambon, 1981; Dierks et al., 1981; Ghosh et al., 1981; Kamen et al., 1982; McKnight and Kingsbury, 1982; Dierks et al., 1983). (Note, however, that it has been recently reported that yeast mRNA initiation sites are determined primarily by specific sequences, not by the distance from the TATA element [Chen and Struhl, 1985; Hahn et al., 1985].)

It was therefore important to determine whether the operator insertions in the plasmids studied here would influence the choice of the transcription start site. We therefore performed CAT-specific primer extension experiments with cytoplasmic RNA extracted from LTK$^-$ cells that had been stably transfected with each MSV–CAT fusion gene tested. (Stable transfectants were used because of the greater specific activity of the CAT RNA in these samples.) We used as primer a 36 nucleotide 5′-end-labeled oligonucleotide complementary to a sequence within the CAT gene (see Experimental Procedures). With RNA from cells transfected with the parental plasmid pSM12, the primer extension product was extended to a length of 144 nucleotides, as expected (Figure 7A, lane 3; the relevant sequences are shown in the legend). In the plasmid pSMAO1, an operator segment of length 40 bp had been substituted for a 17 bp StuI–HindIII fragment between the initiation codon and the transcription start site. With the resulting RNA, the DNA primer was extended to about 168 nucleotides (Figure 7A, lane 4). Thus, the distance between the TATA region and the start site of RNA synthesis in this plasmid was the same as that in the parental plasmid pSM12. Furthermore, Figure 7B shows that mRNA from the plasmid pSMCO1 (lane 2), carrying one operator between the TATA region and the SV40 21 bp repeats, has the same 5′ end as the mRNA from the parental plasmid pSM12. Therefore operator insertions downstream of the transcription start site or upstream of the TATA box do not affect the position of the transcription start site. However, when RNA from cells transfected with pSMBO1, containing one operator insertion between the transcription start site and the TATA box, was used as template, the same primer was extended to 184 nucleotides (Figure 7A, lane 2, and see legend). This corresponds to a new start site at a T residue within the operator sequence. This new start site is 25 bp downstream of the residue T_{15} of the TATA box (numbering as shown in Takahashi et al., 1986). Thus, these results support the view that the TATA region has a "fixing" function for the initiation of transcription in mammalian cells.

Transfected Supercoiled DNA Becomes Nicked in LTK$^-$ Cells but Remains Mostly Intact in LI-1 Cells

We sought to confirm that the apparent repressor-mediated repression of operator-containing MSV–CAT

promoters was not caused by variation in plasmid copy number in the transiently transfected cells. Accordingly, total DNA was isolated from nuclei from transfected LI-1 cells by the Hirt procedure (Hirt, 1967) and was analyzed by electrophoresis on 1% agarose gels and subsequent blot hybridization to a ^{32}P-labeled CAT probe. Results presented in Figure 8 show that the yields of the various CAT plasmid DNAs, as determined by Southern blotting, could not account for the differences in RNA expression or CAT enzyme levels. It was quite surprising that most of the transfected supercoiled DNA remained intact after 48 hr in the repressor-producing cells LI-1. Previous studies have shown that in transfection experiments, extracellular DNA that is not washed out when the culture medium is changed is completely degraded over a 22 hr period following transfection (Alwine, 1985). Therefore the DNA studied here, as isolated by the Hirt (1967) procedure, had been incorporated into the cells. In contrast, in other studies it has been reported that most of the transfected supercoiled DNA becomes nicked or linearized during the course of transfection (Weintraub et al., 1986). To investigate this further, we analyzed DNA from transfected LTK$^-$ cells by Southern blotting as described above. Figure 9 shows that after 48 hr the input supercoiled DNA was largely nicked (27%–62%), but only a small fraction (8%–13%) of DNA became linearized. (Nicked molecules were not distinguished from closed-circular relaxed molecules since they comigrated under the electrophoresis conditions used.) Moreover, some transfected DNA (10%–50%) became concatenated. We note that we did not see concatenation of transfected DNA in LI-1 cells (Figure 8), probably because most of the DNA did not become linearized. Thus there is less nicking and cutting of the plasmid DNA in the repressor-positive LI-1 cells than in the repressor-negative LTK$^-$ cells. This phenomenon does not depend on the presence of an operator sequence in the plasmid; it was observed for pSM12 also. A plausible but speculative interpretation of these results is that the plasmids picked up a protective coat of *lac* repressor protein—because of its nonspecific DNA-binding properties (Lin and Riggs, 1972)—during their passage through the cytoplasm to the nucleus, where they became packaged into chromatin as minichromosomes (Cereghini and Yaniv, 1984; Reeves et al., 1985; Weintraub et al., 1986). Furthermore, transfected DNA was found to be mostly unintegrated in LTK$^-$ and in LI-1 cells in accordance with a previous report (Weintraub et al., 1986). The results presented in Figure 9 also indicate that the quantity of transfected DNA taken up and persisting in nuclei was approximately the same for each plasmid,

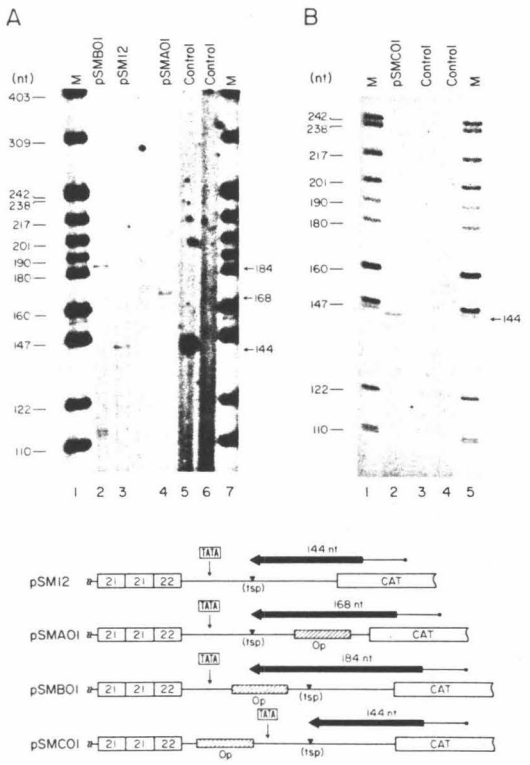
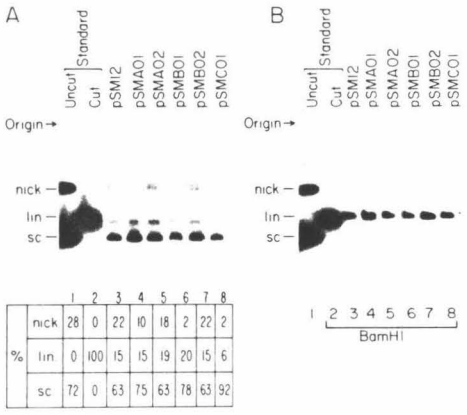


Figure 7. Primer Extension Analysis of Cytoplasmic RNAs Produced by the MSV–CAT Fusion Genes Containing Operator Insertions

(A) The 5′-^{32}P-labeled 36 nucleotide synthetic oligonucleotide complementary to the CAT coding sequence (see Experimental Procedures) was hybridized to approximately 70 μg total cytoplasmic RNA from LTK$^-$ cells transfected with the indicated plasmid constructs. The primer was then extended with unlabeled dNTPs with reverse transcriptase, and the products were fractionated on a 6% polyacrylamide sequencing gel. As control experiments, the same primer was hybridized to either the same amount of total cytoplasmic RNA from nontransfected LTK$^-$ cells (lane 5) or no RNA (lane 6). (B) The same primer was hybridized to approximately 70 μg total cytoplasmic RNA from LTK$^-$ cells transfected with pSMCO1 (lane 2) and then extended as described above. The extended product is displayed on a 6% polyacrylamide sequencing gel. The control experiments were performed as described above (lane 3, RNA from nontransfected LTK$^-$ cells; lane 4, no RNA). ^{32}P-labeled HpaII fragments of pBR322 served as nucleotide size markers (M), which are marked in the left margins. The band due to the extended product is identified by an arrow, and its length is indicated in the right margin of each panel.

Synthesis of the observed cDNA products is depicted schematically below the figure. Symbols are the same as described in the legend to Figure 4. The 40 bp of the *lac* operator insertions is indicated by the hatched box. The primer is shown as a thin line, and extension products as thick arrows. The relevant sequences are shown below:

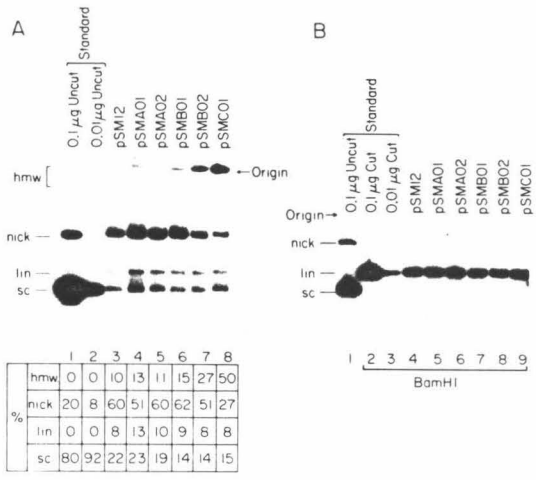
The dashed lines underneath the sequences indicate the extended products, and asterisks denote the sequences deleted in the course of insertion of the *lac* operator sequences.



		1	2	3	4	5	6	7	8
%	nick	28	0	22	10	18	2	22	2
	lin	0	100	15	15	19	20	15	6
	sc	72	0	63	75	63	78	63	92

Figure 8. Southern Blot Analysis of Transfected DNA in the Nuclear Fraction of LI-1 Cells

Total DNA was isolated from nuclei from LI-1 cells transfected 48 hr previously with the indicated plasmids in the absence of IPTG; DNA was analyzed by electrophoresis on 1% agarose gels and subsequent blot hybridization to an SP6-generated CAT probe (from p64-CAT, which gives sense-strand RNA). (A) DNA was undigested; (B) DNA was digested with BamHI prior to electrophoresis. The lanes marked uncut (lane 1) and cut (lane 2) are samples of standard plasmid (0.1 μg of pSM12), either intact or after digestion with BamHI, respectively. Approximately equal amounts of DNA were loaded in each lane. The form of the input DNA is indicated (nick., nicked; lin., linearized; sc., supercoiled), and the relative amount of DNA, as shown below (A), was quantitated by densitometric scanning and then normalized as a percentage of the total CAT DNA.



		1	2	3	4	5	6	7	8
%	hmw	0	0	10	13	11	15	27	50
	nick	20	8	60	51	60	62	51	27
	lin	0	0	8	13	10	9	8	8
	sc	80	92	22	23	19	14	14	15

Figure 9. Topology of Transfected DNA in Isolated Nuclei of LTK⁻ Cells

Total DNA was isolated from nuclei from transfected LTK⁻ cells at 48 hr after transfection and then analyzed and quantitated as described in Figure 8. (A) Uncut DNA; (B) DNA digested with BamHI before being separated on 1% agarose gels. High molecular weight DNA is indicated as hmw. All other symbols are the same as in Figure 8.

and thus could not account for the differences in CAT expression between the operator-containing derivatives and the parental plasmids.

Discussion

Our overall result is that the bacterial *lac* repressor protein encoded in a suitable expression vector can be synthesized in animal cells, assemble into a tetramer, enter the nucleus, and repress expression of another gene that has one or several *lac* operator sequences inserted into any one of several sites in the promoter region of this plasmid. Derepression can be achieved by exposure of the cells to IPTG.

In eukaryotes, most regulatory proteins are transported from their cytoplasmic site of synthesis and are localized primarily in the nucleus. This relatively strict nuclear accumulation is attributed to the presence in the protein of a distinctive stretch of amino acids that acts as a nuclear-localization signal (Goldfarb et al., 1986; Richardson et al., 1986; and also see Dingwall, 1985, for review).

We observe that about 10% of the total *lac* repressor in LI-1 cells, or about 2500 molecules, is located in the nucleus. Since the nuclear volume is about 6% of the total cell volume in a mammalian cell (Alberts et al., 1983), the average concentration of repressor is about the same in the nucleus and cytoplasm. While this result can be rationalized in a number of ways, we find it rather surprising in view of the relatively strong nonspecific DNA-binding properties of repressor. In any case, with a diploid genome size of 6×10^9 bp, LI-1 cells contain approximately one repressor per 2×10^6 bp of DNA. In E. coli, with a genome size of 4.2×10^6 bp and perhaps two chromosomes per cell, and with ca. 10 repressor molecules per cell (Müller-Hill et al., 1968), there is ca. one repressor per 4.2×10^5 bp. These numbers are not greatly different. It is of interest to note that Chao et al. (1980) have reported that, in vitro, assembly of *lac* operator–containing DNA into nucleosomes with octameric histone cores does not affect specific sequence recognition by repressor, or IPTG-induced dissociation, but that nonspecific binding of repressor to DNA is considerably diminished. Specific recognition presumably shows that the repressor-binding surface of the operator DNA faces outward on the nucleosome. Our results showing specific gene regulation by *lac* repressor but limited overall nuclear localization are totally consistent with their results. We note further that Silver et al. (1986) have reported very recently that the DNA-binding protein *lexA* and a *lexA* chimeric DNA regulatory protein are localized predominantly in the cytoplasm in yeast cells.

The data in Figure 8 show that 48 hr after transfection, there are about 150 6 kb plasmid molecules per nucleus, corresponding to an additional 9×10^5 bp. These DNA molecules have entered the nucleus by passage through the cytoplasm. They may have thus acquired an additional coating of repressor, as suggested by the apparent protection against nicking observed in the LI-1 cells. (However, our preliminary results with cells in which some of the plasmids derived here have been stably transfected into the chromosomal DNA of LI-1 cells are that the levels of repression are about the same. Thus, the phenomena

of repression are not dependent on recent exposure of the DNA to the cytoplasm.)

By comparison of the expression levels of the plasmids in repressor-positive LI-1 cells and repressor negative LTK$^-$ cells, we deduce repression levels of 8–12-fold for a single operator insert and 24–48-fold for two in tandem. This is substantially less than the ratio of induced to repressed expression of the *lac* operon in E. coli, which may be as high as 1000-fold (Gilbert and Müller-Hill, 1970). The reasons for this are unknown. They may be related to the different molecular environments of the operator DNA sequence in eukaryotic chromatin and in a bacterial cell. However, our results are concordant with recent experiments that show approximately 10-fold repression, by the mechanism of transcription termination, where the *lac* operator sequence is inserted downstream of coliphage T5 promoter and assayed in E. coli (Deuschle et al., 1986). Whether the extent of repression could be augmented by increasing the overall level of expression of repressor protein or by adding a eukaryotic nuclear localization signal at an appropriate point in the protein is not known at present.

Furthermore, we note that while a significant level of derepression (60%–80%) can be achieved with IPTG, it is not fully reversible, and it occurs slowly (ca. 24 hr). This limited induction may be related to the limited permeability of the eukaryotic plasma membrane to IPTG, but we have no data bearing on this question. In E. coli, IPTG normally enters through the inducible *lac* permease system (Kennedy, 1970); nevertheless, permease-negative mutants are still inducible for *lacZ* by IPTG, although the inducer uptake rate is reduced (Herzenberg, 1959).

From a practical point of view of an inducible genetic switch, the induction ratios achieved at present with the *lac* operator system are not better than those that have been achieved with heat shock, mouse mammary tumor virus, and metallothionein promoters. There may, however, be some situations and some promoters for which the use of the *lac* operator system is advantageous. Further work may reveal methods for achieving higher induction ratios. It will be of interest to test whether the induction ratios observed here can be increased by applying glucocorticoids, since it has been recently reported that the expression of genes fused downstream of the MSV LTR is stimulated 3- to 5-fold by hormone treatment (DeFranco and Yamamoto, 1986).

Our results on the effects of operator insertions and repressor binding on transcript levels have interesting implications for general mechanisms of regulation of gene expression. Specific cases are discussed below. We observe that inserting either one or two copies of the 40 bp operator sequence between the TATA box and the 21 bp repeats reduces expression by a factor of approximately 5 (pSMCO1 and pSMCO2, Figure 4). Specific transcription factors related to the SP1 factor are believed to bind to the 21 bp repeats and contact a protein bound to the TATA box (for review, see Dynan and Tjian, 1985); the latter, in turn, interacts with RNA polymerase II to activate the transcriptional machinery. When this SP1 factor and the TATA box factor are separated by one or two *lac* operators, transcription rates are reduced. The interactions must now require interaction at a distance—for example, by looping (Ptashne, 1986). Binding of repressor to the operator sequence in this position greatly inhibits transcription; presumably the *lac* repressor protein presents a steric block to the normal protein–protein interactions.

Takahashi et al. (1986) have tested the effects of altering the stereospecific relationships of SV40 promoter elements by constructing insertions of 4 or 15 bp between the 21 bp repeats and the TATA box region. These insertions decrease transcription in vivo more drastically than insertions of 10 or 21 bp because they place the interacting proteins on opposite sides of the DNA helix. The *lac* operator sequence used by us is 40 bp in length, which is approximately an integral number of turns. Thus the effects observed here in the absence of repressor are mainly due to increased distance, not opposite orientation relative to the helix axis.

In agreement with earlier studies, we observed that placement of a 40 bp operator segment between the TATA box and the normal transcription start point changes the start point to a new site, which is again about 25 nucleotides from the TATA box (see legend to Figure 7). In the presence of the repressor, transcription is reduced by a factor of about 12. We were unable to map the start point in the repressed state because of the amount of RNA produced; if it is the same, we assume that the binding of a protein to the TATA box or of RNA polymerase II close to the transcription start point can displace the binding of repressor to the operator.

Finally, binding of repressor to an operator inserted downstream of the transcription start point also inhibits transcription. This is most plausibly interpreted as a block to passage of the transcription complex, in agreement with recent results for E. coli RNA polymerase in vivo and in vitro (Deuschle et al., 1986). In addition, we note that insertion of operator sequences between the transcription start point and the initiation codon may result in the formation of a hairpin structure in the 5′-untranslated region of the mRNA, thus affecting translation initiation. The *lac* operator sequence has extensive bilateral symmetry (Barkley and Bourgeois, 1980). This hypothesis is supported by recent experiments in which hairpin structures were generated upstream of the initiator codon by oligonucleotide insertion, thus leading to a drastic reduction in translation (Kozak, 1986).

Experimental Procedures

Site-Directed Mutagenesis of the *lacI* Gene

The *lacI*^{q} gene contained in plasmid pIQ (J. L. Betz, unpublished results) was excised by EcoRI and HincII digestions, and subcloned into the EcoRI and HincII sites of the vector M13mp19 (Yanisch-Perron et al., 1985). The oligonucleotide 5′-TGGTGAATATGAAACCAG-3′ was employed as a mismatched primer to introduce a G-to-A mutation of the GTG initiator codon. The 5′-phosphorylated mutagenic primer and the M13 primer 5′-ATGCCCCCTGCCTATTTCGG-3′ (nucleotides 2054–2073 of M13mp19) were simultaneously annealed to a single-stranded noncoding template of the *lacI*-M13 clone, extended with E. coli DNA polymerase I large fragment (Klenow), and ligated essentially as described by Zoller and Smith (1984). Following transformation into strain JM101 and screening for the mutant template, single-stranded DNA of the mutated *lacI* (*lacI*^{m})–M13 was prepared, and the

entire DNA insert was sequenced by the dideoxy chain-terminating method (Sanger et al., 1977; Biggin et al., 1983). Double-stranded (M13 RI form) DNA for the mutant *lacI*^{m} gene was excised with EcoRI and HincII and was reintroduced into the corresponding sites of an SP62 vector together with the 0.8 kb HincII–EcoRI DNA fragment of the 3′ portion of the *lacI* gene. This mutant plasmid was designated pI^{m}SP62. To delete the operator sequence downstream of the termination codon of the *lacI* gene, pI^{m}SP62 DNA was digested with BamHI and SphI, thus generating a 3′-protruding end that protected the remainder of the vector from exonuclease III attack, and unidirectionally deleted by exonuclease III (Henikoff, 1984). This DNA was rendered blunt-ended with T4 DNA polymerase, ligated with T4 DNA ligase, and used to transform the MC1061.2 strain of E. coli. Subsequently, these plasmids were excised by EcoRI and SalI digestions, and the mutant *lacI*^{m}-containing fragments were cloned into the corresponding sites of the M13mp18 vector. Single-stranded DNA was prepared from a plaque-purified phage, and the DNA sequence at the deletion junction was determined using the dideoxy chain-terminating method (Sanger et al., 1977; Biggin et al., 1983).

Construction of the *lacI* Expression Vector pRSV-I

The plasmid pI^{m}SP62, carrying the entire sequence of the *lacI* gene (from 30 nucleotides upstream of the primary translation initiation site to 18 nucleotides downstream of the termination codon) was digested with EcoRI, blunt-ended with Klenow fragment in the presence of dNTPs, annealed with HindIII linker, ligated with T4 DNA ligase, and digested with HindIII. The 1.3 kb HindIII–SalI fragment was purified by electrophoresis in low-gelling-temperature agarose (1.0% SeaPlaque agarose; FMC Corporation, Marine Colloids Division) and was subcloned into pRSV-CAT (Gorman et al., 1982b). This plasmid was digested with HpaI, annealed with SalI linker, ligated with T4 DNA ligase, and then redigested with SalI and HindIII. The 3.5 kb HindIII–SalI fragment was gel-isolated and then ligated with the 1.3 kb HindIII–SalI fragment of *lacI*^{m} and then used to transform MC1061.2. In the resulting plasmid, designated pRSV-I, a *lacI*-encoding region is installed under transcriptional control of the RSV LTR. This plasmid carried an SV40 poly(A) addition signal and was used for the expression of the *lac* repressor protein in LTK$^-$ cells (see below).

Construction of CAT Plasmids Containing Operator Inserts

Plasmid pOE101, carrying 12 tandem copies of the *lac* operator (Sadler et al., 1980), was partially digested with EcoRI. The ends of the operator monomer and oligomers were then repaired with Klenow fragment in the presence of dNTPs. Operator-inserted CAT plasmids were constructed by the following procedures. First, for insertion between the initiation codon (ATG) and the transcription start point, parental plasmids pSM11 and pSM12 (containing the MSV enhancer in the opposite orientations, respectively, and also containing the SV40 21 bp repeats and its promoter, the coding sequences from the CAT gene, and the SV40 poly[A] addition signal; Laimins et al., 1984) were linearized by digestion with StuI and HindIII. The DNA molecules were fractionated by agarose gel electrophoresis, and the linearized large fragments (about 6 kb) of the plasmids were extracted and purified. The ends of the DNA fragments were repaired with Klenow fragment in the presence of dNTPs, and the blunt-ended operator monomer and oligomers mixture was then added and ligated with T4 DNA ligase. Thus, the StuI–HindIII segment (17 bp) of each parental plasmid was replaced by single or multiple tandem operator sequences in the resultant plasmids. Second, for insertion between the transcription start point and the TATA box region, pSM11 and pSM12 was digested with BglI, repaired with Klenow fragment in the presence of dNTPs, incubated with the blunt-ended operator monomer and oligomer mixture, and simultaneously ligated with T4 DNA ligase and used to transform MC1061.2. The resultant plasmids thus contained single or multiple tandem operator sequences between the transcription start and the TATA box regions. Third, for insertion between the TATA box and SV40 21 bp repeats regions, pSM11 or pSM12 was partially digested with NcoI, and the full-length linearized plasmid DNA was fractionated by agarose gel electrophoresis. The ends of the DNA were filled in with Klenow fragment in the presence of dNTPs, the blunt-ended operator monomer and oligomers were added, and the DNA mixture was ligated with T4 DNA ligase. The resultant plasmids therefore included single or multiple tandem operator sequences between the TATA box and the SV40 21 bp repeat regions. Subsequently, all of the operator-containing plasmids were excised by EcoRI and SalI digestion and were cloned into the corresponding sites of M13mp18. That single or multiple tandem operator sequences were in fact introduced into the CAT plasmids at the sites indicated was verified by DNA sequencing (Sanger et al., 1977; Biggin et al., 1983).

Cell Culture and Stable Transfectants

Mouse LTK$^-$ cells were grown in Dulbecco's modified Eagle's (DME) medium supplemented with 10% fetal calf serum. Cells to be transfected were plated on the day before transfection at a density of 1.0×10^6 cells per 60 mM dish. LTK$^-$ cells were cotransfected with a 50:1 molar ratio of pRSV-I and pTK-5 (Roach et al., 1984), which carries the 3.5-kb *tk*-coding BamHI fragment of the herpes simplex virus type I *tk* gene. Calcium phosphate precipitates of the plasmid DNA were prepared as described elsewhere (Gorman et al., 1982a). The precipitate was left on the cells for 16 hr, a fresh aliquot of medium was supplied, and the cells were incubated without selection for 24 hr prior to addition of HAT medium (DME medium containing 10% fetal calf serum, 0.1 mM hypoxanthine, 0.4 μM aminopterin, 16.0 μM thymidine). Resistant colonies were isolated and expanded after 14 days. Transfectants expressing the *lacI* gene were initially identified by RNA dot blots and were then characterized by an operator-binding assay to quantitate the amount of functional repressor protein (see below). In addition, the levels of repressor protein in the nucleus and in the cytoplasm were quantitated and estimated by Western blotting and radiographic detection with antibody.

Transient Transfection and CAT Assay

LTK$^-$ cells and *lac* repressor–producing cells (LI-1) were grown in DME supplemented with 10% fetal calf serum and in HAT medium, respectively. Cells to be transfected were plated on the day prior to transfection at a density of 1.2×10^6 cells per 60 mM dish. DEAE-dextran transfections were performed by washing the cells twice with DME without serum and then adding 500 μg/ml DEAE-dextran (Sigma Chemical Co., MW 500,000) in DME with the appropriate DNA in the presence of chloroquine (2 μg/ml). Incubations were for 4 hr at 37°C including 1 hr of initial incubation in 2% CO_2 and 3 hr of incubation in 5% CO_2 as described by Sussman and Milman (1984). The cells were shocked with dimethyl sulfoxide for 2 min at 37°C according to Sussman and Milman (1984); the media were replaced, either with or without IPTG (20 mM final concentration), and the cells were incubated for 48 hr before harvesting. Each plasmid preparation used for transfection was purified through two CsCl-ethidium bromide equilibrium gradients. After transfection, the CAT extractions and assay were performed using equal amounts of protein from the cellular extracts according to Gorman et al. (1982a).

Isolation and Analysis of Plasmid DNA from Transfected Cells

Total DNA was isolated from nuclei from transfected cells by the method of Hirt (1967). The supernatants were phenol-extracted twice, chloroform-extracted once, and the DNA was ethanol-precipitated. The pellets were washed with 75% ethanol, dried, and dissolved in 20 μl of TE (10 mM Tris–HCl, 1 mM EDTA). Equivalent portions of each sample were digested with BamHI. The digests and an equivalent portion of each undigested sample were electrophoresed on 1% agarose gels and then blotted to Zeta-probe (Bio-Rad) or Hybond-N (Amersham) membranes, and plasmid DNA was visualized by hybridization with ^{32}P-labeled SP6-CAT RNA. After autoradiography, the resultant bands were quantitated by densitometric scanning and comparison with the intensities of bands of known amounts of purified plasmid DNA.

Isolation and Analysis of RNA from Transfected Cells

Cytoplasmic RNA was isolated from mouse LTK$^-$ or LI-1 cells. Transfected cells were washed with PBS (phosphate-buffered saline), scraped off with 5 ml of PBS, and centrifuged. The pellet was resuspended in 0.1 ml of 0.25 M Tris–HCl, and the cell suspension was frozen and quickly rethawed four times on dry ice. Nuclei were removed by centrifugation in an Eppendorf microfuge for 15 min (4°C), and RNA was isolated from the resulting supernatant according to Gorman et al. (1982b). An equal volume of a solution of 7 M urea, 0.35 M NaCl, 10 mM Tris–HCl (pH 8.0), 10 mM EDTA, 1% SDS and about 40 μg of tRNA per ml were added. The mixture was extracted twice with

phenol–chloroform–isoamyl alcohol (20:20:1 [v/v]) and twice with chloroform–isoamyl alcohol. After precipitation with ethanol the RNA was collected by centrifugation and was resuspended in 10 mM Tris–HCl, 0.1 mM EDTA, 0.1% SDS. Equivalent portions of each RNA sample were assayed by dot-blot analysis using an SP6-generated antisense *lacI* probe, an SP6-generated antisense CAT probe, or an SP6-generated antisense actin probe (Hu et al., 1986). Approximately 70 μg of each cytoplasmic RNA preparation was used for each primer extension experiment, and the 36 nucleotide ^{32}P-labeled oligonucleotide 5′-TGCCATTGGGATATATCAACGGTGGTATATCCAGTG-3′ (complementary to a region of the CAT mRNA started at 12 nucleotides downstream of the initiation codon AUG) was used as primer. Primer extension was performed as previously described (Hu et al., 1986) except that the cloned Moloney murine leukemia virus reverse transcriptase (BRL) was used in extension.

Western Blot Analysis

Crude cytoplasmic and nuclear extracts from LI-1 and LTK$^-$ cells were prepared from two 100 mM dishes near confluence. The cells were washed with PBS, scraped off with 5 ml of PBS, and centrifuged. The pellet was resuspended in 0.1 ml of cold lysis buffer containing 0.2 M potassium phosphate (K_2HPO_4:KH_2PO_4 = 5:1, pH 7.4), 0.1 mM EDTA, 0.3 mM DTT, 5% (w/v) glucose, 25 μM leupeptin, 1 μg/ml bestatin, 10 μg/ml trypsin inhibitor, and 1 mM PMSF. The cell suspension was freeze-thawed four times on dry ice, and the nuclei and the cell debris were sedimented. The supernatant was used for Western blot analysis and for the repressor–operator filter-binding assay (see below). The nuclear pellet was resuspended in 0.1 ml of homogenization buffer (0.5% NP40, 0.2 mM EDTA) and homogenized. The homogenate was treated with DNAase I (1 mg/ml final concentration) for 1 hr at 37°C and then mixed with the electrophoresis sample solution (2× ESS: 20 mM Tris–HCl, 5 mM EDTA, 4% SDS, 20% glycerol, 10% β-mercaptoethanol, 0.5 mg/ml bromophenol blue [pH 7.0]). Equal amounts of protein from cytoplasmic extracts of LI-1 and LTK$^-$ cells (~2 × 10^6) were also mixed with equal volumes of 2× ESS. In addition, ^{14}C-labeled molecular weight standards (BRL) were included. The samples were subjected to SDS–10% polyacrylamide gel electrophoresis and then transferred to nitrocellulose according to Burnette (1981). The filter was probed with anti-repressor monoclonal antibody B-2 (Sams et al., 1985) followed by ^{125}I-labeled sheep anti-mouse immunoglobulin (Amersham), air-dried, and exposed to X-ray film. The amount of repressor protein was quantitated by scanning the autoradiograph.

Other Methods

Nucleic acid manipulations and blot hybridizations were according to standard procedures (Maniatis et al., 1982; Reed and Mann, 1985). In vitro transcription with SP6 polymerase employed methods and reagents as described by Melton et al. (1984). The filter-binding assay of repressor–operator complexes was performed essentially as described by Lin and Riggs (1972) except that the DNA probe used was a gel-purified ^{32}P-end-labeled operator fragment of 40 bp. Protein content of cellular extracts was measured by the method of Bradford (1976).

Acknowledgments

We are grateful to Joan Betz for generously providing plasmids pIQ and pOE101, to George Khoury for plasmids pSM11 and pSM12, to Kathleen Matthews for the monoclonal antibody to the *lac* repressor, and to Sheldon York for the purified E. coli *lac* repressor. We thank Marie Krempin for her assistance with cell cultures. We thank Nevis Fregien, Carl Parker, Terry Snutch, and Barbara Wold for helpful discussions and comments on this work and manuscript. This work was supported by a Public Health Service research grant from the National Institutes of Health.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received November 12, 1986; revised December 16, 1986.

References

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., and Watson, J. D. (1983). Molecular Biology of the Cell. (New York: Garland Publishing), p. 321.

Alwine, J. C. (1985). Transient gene expression control: effects of transfected DNA stability and trans-activation by viral early proteins. Mol. Cell. Biol. *5*, 1034–1042.

Barkley, M. D., and Bourgeois, S. (1980). Repressor recognition of operator and effectors. In The Operon, J. H. Miller and W. S. Reznikoff, eds. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory), pp. 177–220.

Beckwith, J. R., and Zipser, D. (1970). The Lactose Operon. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory).

Benoist, C., and Chambon, P. (1981). *In vivo* sequence requirements for the SV40 early promoter region. Nature *290*, 304–310.

Biggin, M. D., Gibson, T. J., and Hong, G. F. (1983). Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. Proc. Natl. Acad. Sci. USA *80*, 3963–3965.

Bradford, M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein dye binding. Anal. Biochem. *72*, 248–254.

Brent, R., and Ptashne, M. (1984). A bacterial repressor protein or a yeast transcriptional terminator can block upstream activation of a yeast gene. Nature *312*, 612–615.

Burnette, W. N. (1981). "Western blotting": electrophoretic transfer of proteins from sodium dodecyl sulfate-polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radioiodinated protein A. Anal. Biochem. *112*, 195–203.

Cereghini, S., and Yaniv, M. (1984). Assembly of transfected DNA into chromatin. EMBO J. *3*, 1243–1253.

Chao, M. V., Gralla, J. D., and Martinson, H. G. (1980). *lac* operator nucleosomes. I. Repressor binds specifically to operator within the nucleosome core. Biochemistry *19*, 3254–3260.

Chen, W., and Struhl, K. (1985). Yeast mRNA initiation sites are determined primarily by specific sequences, not by the distance from the TATA element. EMBO J. *4*, 3273–3280.

Davidson, I., Fromental, C., Augereau, P., Wildeman, A., Zenke, M., and Chambon, P. (1986). Cell-type specific protein binding to the enhancer of simian virus 40 in nuclear extracts. Nature *323*, 544–548.

DeFranco, D., and Yamamoto, K. R. (1986). Two different factors act separately or together to specify functionally distinct activities at a single transcriptional enhancer. Mol. Cell. Biol. *6*, 993–1001.

Deuschle, U., Gentz, R., and Bujard, H. (1986). *lac* repressor blocks transcribing RNA polymerase and terminates transcription. Proc. Natl. Acad. Sci. USA *83*, 4134–4137.

Dierks, P., van Ooyen, A., Mantei, N., and Weissmann, C. (1981). DNA sequences preceding the rabbit β-globin gene are required for formation in mouse L cells of β-globin RNA with the correct 5′ terminus. Proc. Natl. Acad. Sci. USA *78*, 1411–1415.

Dierks, P., van Ooyen, A., Cochran, M. D., Dobkin, C., Reiser, J., and Weissmann. C. (1983). Three regions upstream from the cap site are required for efficient and accurate transcription of the rabbit β-globin gene in mouse 3T6 cells. Cell *32*, 695–706.

Dingwall, C. (1985). The accumulation of proteins in the nucleus. Trends Biochem. Sci. *10*, 64–66.

Dynan, W., and Tjian, R. (1985). Control of eukaryotic messenger RNA synthesis by sequence-specific DNA-binding protein. Nature *316*, 774–778.

Fregien, N., and Davidson, N. (1986). Activating elements in the promoter region of the chicken β-actin gene. Gene *46*, in press.

Ghosh, P. K., Lebowitz, P., Frisque, R. J., and Gluzman, Y. (1981). Identification of a promoter component involved in positioning the 5′ termini of simian virus 40 early mRNAs. Proc. Natl. Acad. Sci. USA *78*, 100–104.

Gilbert, W., and Müller-Hill, B. (1970). The lactose repressor. In The Lactose Operon, J. R. Beckwith and D. Zipser, eds. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory), pp. 93–109.

Goldfarb, D. S., Gariepy, J., Schoolnik, G., and Kernberg, R. D. (1986). Synthetic peptides as nuclear localization signals. Nature *322*, 641–644.

Goodbourn, S., Zinn, K., and Maniatis, T. (1985). Human β-interferon gene expression is regulated by an inducible enhancer element. Cell *41*, 509–520.

Gorman, C. M., Moffat, L. F., and Howard, B. H. (1982a). Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. Mol. Cell. Biol. *2*, 1044–1051.

Gorman, C. M., Merlino, G. T., Willingham, M. C., Pastan, I., and Howard, B. H. (1982b). The Rous sarcoma virus long terminal repeat is a strong promoter when introduced into a variety of eukaryotic cells by DNA-mediated transfection. Proc. Natl. Acad. Sci. USA *79*, 6777–6781.

Grosschedl, R., and Birnstiel, M. L. (1980). Identification of regulatory sequences in the prelude sequences of an H2A histone gene by the study of specific deletion mutants *in vivo*. Proc. Natl. Acad. Sci. USA *77*, 1432–1436.

Hahn, S., Hoar, E. T., and Guarente, L. (1985). Each of three "TATA elements" specifies a subset of the transcription initiation sites at the CYC-1 promoter of *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA *82*, 8562–8566.

Henikoff, S. (1984). Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene *28*, 351–359.

Herzenberg, L. A. (1959). Studies on the induction of β-galactosidase in a cryptic strain of *Escherichia coli*. Biochim. Biophys. Acta *31*, 525–538.

Hirt, B. (1967). Selective extraction of polyoma DNA from infected mouse cell cultures. J. Mol. Biol. *26*, 365–369.

Hu, M. C.-T., Sharp, S. B., and Davidson, N. (1986). The complete sequence of the mouse skeletal α-actin gene reveals several conserved and inverted repeat sequences outside of the protein-coding region. Mol. Cell. Biol. *6*, 15–25.

Kamen, R., Jat, P., Treisman, R., Favaloro, J., and Folk, W. R. (1982). 5′ Termini of polyoma virus early region transcripts synthesized *in vivo* by wild-type virus and viable deletion mutants. J. Mol. Biol. *159*, 189–224.

Kennedy, E. P. (1970). The lactose permease system of *Escherichia coli*. In The Lactose Operon, J. R. Beckwith and D. Zipser, eds. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory), pp. 49–92.

Kozak, M. (1986). Influence of mRNA secondary structure on initiation by eukaryotic ribosomes. Proc. Natl. Acad. Sci. USA *83*, 2850–2854.

Laimins, L. A., Gruss, P., Pozzatti, R., and Khoury, G. (1984). Characterization of enhancer elements in the long terminal repeat of Moloney murine sarcoma virus. J. Virol. *49*, 183–189.

Lin, S., and Riggs, A. D. (1972). *lac* repressor binding to non-operator DNA: detailed studies and a comparison of equilibrium and rate competition methods. J. Mol. Biol. *72*, 671–690.

Looman, A. C., and van Knippenberg, P. H. (1986). Effects of GUG and AUG initiation codons on the expression of *lacZ* in *Escherichia coli*. FEBS Lett. *197*, 315–320.

Lopata, M. A., Cleveland, D. W., and Sollner-Webb, B. (1986). RNA polymerase specificity of mRNA production and enhancer action. Proc. Natl. Acad. Sci. USA *83*, 6677–6681.

Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982). Molecular Cloning: A Laboratory Manual. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory).

Mathis, D. J., and Chambon, P. (1981). The SV40 early region TATA box is required for accurate *in vitro* initiation of transcription. Nature *290*, 310–315.

Mayo, K. E., Warren, R., and Palmiter, R. D. (1982). The mouse metallothionein-I gene is transcriptionally regulated by cadmium following transfection into human or mouse cells. Cell *29*, 99–108.

McKnight, S. L., and Kingsbury, R. (1982). Transcriptional control signals of a eukaryotic protein-coding gene. Science *217*, 316–324.

McKnight, S., and Tjian, R. (1986). Transcriptional selectivity of viral genes in mammalian cells. Cell *46*, 795–805.

Melton, D. A., Krieg, P. A., Rebagliati, M. R., Maniatis, T., Zinn, K., and Green, M. R. (1984). Efficient *in vitro* synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. Nucl. Acids Res. *12*, 7035–7056.

Miller, J. H., and Reznikoff, W. A. (1980). The Operon. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory).

Müller-Hill, B., Crapo, L., and Gilbert, W. (1968). Mutants that make more *lac* repressor. Proc. Natl. Acad. Sci. USA *59*, 1259–1264.

Ptashne, M. (1986). Gene regulation by proteins acting nearby and at a distance. Nature *322*, 697–701.

Reed, K. C., and Mann, D. A. (1985). Rapid transfer of DNA from agarose gels to nylon membranes. Nucl. Acids Res. *12*, 7207–7221.

Reeves, R., Gorman, C., and Howard, B. (1985). Minichromosome assembly of nonintegrated DNA transfected into mammalian cells. Nucl. Acids Res. *13*, 3599–3615.

Richardson, W. D., Roberts, B. L., and Smith, A. E. (1986). Nuclear location signals in polyoma virus large-T. Cell *44*, 77–85.

Riggs, A. D., Newby, R. F., and Bourgeois, S. (1970). *lac* repressor-operator interaction. II. Effect of galactosides and other ligands. J. Mol. Biol. *51*, 303–314.

Ringold, G. M. (1983). Regulation of mouse mammary tumor virus gene regulation by glucocorticoid hormones. Curr. Topics Microbiol. Immunol. *106*, 79–103.

Roach, A., Nicolson, M. O., and Davidson, N. (1984). Transcription and expression of the herpes simplex virus *tk* gene inserted into proviral sequences of feline leukemia virus. Gene *32*, 389–398.

Ryals, J., Dierks, P., Ragg, H., and Weissmann, C. (1985). A 46-nucleotide promoter segment from an IFN-α gene renders an unrelated promoter inducible by virus. Cell *41*, 497–507.

Sadler, J. R., Tecklenburg, M., and Betz, J. L. (1980). Plasmids containing many tandem copies of a synthetic lactose operator. Gene *8*, 279–300.

Sams, C. F., Hemelt, V. B., Pinkerton, F. D., Schroepfer, G. J., Jr., and Matthews, K. S. (1985). Exposure of antigenic sites during immunization. Monoclonal antibodies to monomer of lactose repressor protein. J. Biol. Chem. *260*, 1185–1190.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA *74*, 5463–5467.

Silver, P. A., Brent, R., and Ptashne, M. (1986). DNA binding is not sufficient for nuclear localization of regulatory proteins in *Saccharomyces cerevisiae*. Mol. Cell. Biol. *6*, 4763–4766.

Steege, D. A. (1977). 5′-Terminal nucleotide sequence of *Escherichia coli* lactose repressor mRNA: features of translational initiation and reinitiation sites. Proc. Natl. Acad. Sci. USA *74*, 4163–4167.

Sussman, D. J., and Milman, G. (1984). Short-term, high-efficiency expression of transfected DNA. Mol. Cell. Biol. *4*, 1641–1643.

Szybalska, E. H., and Szybalski, W. (1962). Genetics of human cell lines. IV. DNA-mediated heritable transformation of a biochemical trait. Proc. Natl. Acad. Sci. USA *48*, 2026–2034.

Takahashi, K., Vigneron, M., Matthes, H., Wildeman, A., Zenke, M., and Chambon, P. (1986). Requirement of stereospecific alignments for initiation from the simian virus 40 early promoter. Nature *319*, 121–126.

Topol, J., Ruden, D. M., and Parker, C. S. (1985). Sequences required for in vitro transcriptional activation of a Drosophila hsp 70 gene. Cell *42*, 527–537.

Weintraub, H., Cheng, P. F., and Conrad, K. (1986). Expression of transfected DNA depends on DNA topology. Cell *46*, 115–122.

Wigler, M., Sweet, R., Sim, G. K., Wold, B., Pellicer, A., Lacy, E., Maniatis, T., Silverstein, S., and Axel, R. (1979). Transformation of mammalian cells with genes from procaryotes and eucaryotes. Cell *16*, 777–785.

Wu, C. (1984). Two protein-binding sites in chromatin implicated in the activation of heat-shock genes. Nature *309*, 229–234.

Yanisch-Perron, C., Vieria, J., and Messing, J. (1985). Improved M13 phage cloning, vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene *33*, 103–119.

Zoller, M., and Smith, M. (1984). Oligonucleotide-directed mutagenesis: a simple method using two oligonucleotide primers and a single-stranded DNA template. DNA *3*, 479–488.

CHAPTER 3

The inducible *lac* operator-repressor system is functional for control of expression of injected DNA in *Xenopus* oocytes.

Mickey C-T. Hu and Norman Davidson

Gene 62:301-313 (1988)

GEN 02257

The inducible *lac* operator-repressor system is functional for control of expression of injected DNA in *Xenopus* oocytes

(Recombinant DNA; transcription repression; transcription induction; genetic switch; microinjection)

Mickey C.-T. Hu and Norman Davidson

Church Chemical Laboratories, 164-30, California Institute of Technology, Pasadena, CA 91125 (U.S.A.)

Received 23 September 1987
Accepted 14 October 1987
Received by publisher 26 October 1987

SUMMARY

We have investigated the use of the *Escherichia coli lac* operator-repressor system to regulate the expression of genes introduced by microinjection into *Xenopus laevis* oocytes. We observe that expression of an MSV-*cat* fusion gene, in which the *lac* operator was inserted between the TATA box and the transcription start point (tsp), or between the tsp and the start codon (ATG), is completely repressed when the *lac* repressor protein is added to the plasmid suspension prior to injection. The *lac* repressor had no detectable effect on the expression of a coinjected HSV-1 *tk* gene that had no operator insertion (or on an MSV-*cat* gene without an operator), indicating that the nonspecific DNA-binding properties of the repressor do not inhibit transcription. CAT activity expressed from the operator-containing MSV-*cat* genes transcribed in the oocyte nucleus was also inhibited by repressor injected into the oocyte cytoplasm, showing that biologically active repressor proteins can enter the nucleus from the cytoplasm. Injection of the inducer IPTG into the oocyte cytoplasm markedly derepressed the repressed *cat* genes but not the HSV-1 *tk* gene coinjected as an internal control. Overall, our results show that the *lac* operator-repressor system can be useful as a genetic switch in the regulation of gene expression of injected DNA in frog oocytes. Finally, our observations on the vectors used in this work show that the MSV enhancer significantly activates transcription from the SV40 early promoter in frog oocytes, although previous studies have indicated that the MSV enhancer is not necessary for the activity of the MSV promoter in oocytes [Graves et al., Mol. Cell. Biol. 5 (1985) 1945–1958].

Correspondence to: Dr. N. Davidson, Church Chemical Laboratories, 164-30, California Institute of Technology, Pasadena, CA 91125 (U.S.A.) Tel. (818)356-6055.

Abbreviations: bp, base pair(s); BSA, bovine serum albumin; buffers B, R, ND-96, see MATERIALS AND METHODS, section **b**: CAT, Cm acetyltransferase; *cat*, gene coding for CAT; Cm, chloramphenicol; dNTP, deoxynucleoside triphosphate; DTT, dithiothreitol; HSV, Herpes simplex virus; IPTG, isopropyl-β-D-thiogalactopyranoside; LTR, long terminal repeat; MSV, Moloney murine sarcoma virus; nt, nucleotide(s); *ori*, origin of DNA replication; PA, polyacrylamide; RSV, Rous sarcoma virus; SDS, sodium dodecyl sulfate; SV, simian virus; *tk*, thymidine kinase gene; tsp, transcription start point.

INTRODUCTION

The control of gene expression can occur, in principle, at a number of levels, the most basic of which is transcription initiation. Repression by binding of the *lac* repressor to the *lac* operator and derepression by action of inducers such as IPTG is one of the most thoroughly studied and best understood examples of a protein–nucleic acid interaction that regulates transcription of a gene in prokaryotic systems (for reviews, see Beckwith and Zipser, 1970; Miller and Reznikoff, 1980). Recently, we and others have reported that the *lac* operator-repressor system is functional in mammalian cells in culture and can be used as an inducible genetic switch for gene transfer experiments (Hu and Davidson, 1987; Brown et al., 1987).

We report here the results of our concurrent studies of the properties of the *lac* operator-repressor system for control of the expression of genes introduced by microinjection into *Xenopus* oocytes. The oocyte system has certain advantages over, and certain differences from, mammalian cells for studies of problems in gene expression. (For general reviews of expression of injected genes in *Xenopus* oocytes, see Gurdon and Melton, 1981; Gurdon and Wickens, 1983; Colman, 1984a.) First, in oocytes, genes can be introduced in controlled quantities and in a controlled physical state (e.g., supercoiled, linearized, *etc.*) by microinjection. By contrast, in transient expression gene transfer experiments in mammalian cells, the copy number of transfected genes per cell is greatly variable with time and quite heterogenous, and the level of expression is quite variable from cell to cell (Domen et al., 1986; Alwine, 1985; Spandidos and Wilkie, 1984). Second, regulatory molecules can be introduced into an oocyte in a controlled manner, but there is no convenient controllable way of introducing a defined amount of a regulatory protein or an inducer into a mammalian cell. Third, the effects of transcriptional enhancers on gene expression are sometimes different in oocytes than in mammalian cells. We anticipate that the present system will be useful as a tool for the study of the mechanism of enhancer function and hence of these differences.

MATERIALS AND METHODS

(a) Plasmid constructions

Plasmid constructions of operator-carrying *cat* plasmids have been described previously (Hu and Davidson, 1987). Plasmid pSMEO1 was constructed by inserting one operator sequence (40 bp) into the *Bgl*I site of pSM11, between the tsp and the TATA box region. Plasmid pSMDO1 was constructed by replacing the *Stu*I-*Hin*dIII segment (21 bp), between the tsp and the start codon (ATG), by one operator sequence (40 bp). Plasmid pTK-5 (Roach et al., 1984) was constructed by inserting the 3.5-kb *Bam*HI fragment containing the *tk* gene of HSV-1 (McKnight, 1980; Wagner et al., 1981) into the *Bam*HI site of pBR322.

(b) Oocyte preparation and injection

Adult female *Xenopus laevis* were purchased from Nasco (Ft. Atkinson, WI), anesthetized by immersion in MS-222 (Sigma), and an ovarian fragment was excised as described (White et al., 1985; Colman, 1984b). Stage-V and stage-VI oocytes (Dumont, 1972) were isolated from the ovarian tissue by collagenase treatment (White et al., 1985) and placed in ND-96 buffer (96 mM NaCl, 2 mM KCl, 1.8 mM $CaCl_2$, 1 mM $MgCl_2$, 5 mM Hepes/NaOH, pH 7.6) supplemented with 2.5 mM sodium pyruvate at 20°C. In preparation for injection, CsCl-density-gradient-purified plasmid DNA was dissolved at a concentration of 350 μg/ml in B buffer (10 mM Tris · HCl, pH 7.4, 3 mM magnesium acetate, 0.1 mM EDTA, 0.1 mM DTT, 50 μg/ml BSA). About 10–20 nl of this solution was injected into the nucleus of each oocyte as described by Gurdon and Wickens (1983) and Colman (1984a). The purified *E. coli lac* repressor (kindly provided by Sheldon York, University of Colorado) was dissolved in R buffer (0.2 M potassium phosphate, pH 7.4, 0.1 mM EDTA, 0.3 mM DTT, 5% [w/v] glucose, 50 μg/ml PMSF). The repressor solution was diluted into B buffer at a final concentration of 0.14 mg/ml prior to injection into the oocyte nucleus. In experiments where the repressor was injected into the cytoplasm, about 50 nl of the repressor solution (1.35 mg/ml) was injected into the cytoplasm of each oocyte. When it was necessary to give oocytes a second

injection in a specific location with respect to the first, only oocytes showing a trace of the first injection site were used (the injection site is very frequently marked by a concentration of pigment). In some experiments where the inducer IPTG was injected, about 20 nl of 0.5 M IPTG (in sterile water) solution was injected into the cytoplasm of each oocyte. Groups of 30 to 60 oocytes were injected in each experiment and then incubated in ND-96 buffer supplemented with 10 μg/ml penicillin and 10 μg/ml streptomycin for 48 h at 20°C. Batches of ten healthy-looking oocytes were taken for analysis of RNA or enzyme activity.

(c) Extract preparation and CAT assay

Microinjected oocytes were washed briefly in ND-96 buffer, suspended in 200 μl of homogenization buffer (100 mM NaCl, 20 mM Tris · HCl, pH 7.6, 1% Triton X-100, 1 mM PMSF), homogenized and centrifuged (12000 × *g*, 15 min) in a microcentrifuge to pellet insoluble material. The cytoplasmic extract was then removed carefully to avoid lipid contamination, and a 20-μl portion was incubated with a solution of 100 μl of 250 mM Tris · HCl (pH 7.8), 20 μl of 4 mM acetyl coenzyme A, and 10 μl of ^{14}C-labeled Cm (0.2 μCi, approx. 50 mCi/mmol; Amersham Corp.) for 2 h at 37°C. *E. coli* CAT (P-L Biochemicals Inc.) was used as a control. Cm was then extracted with 1 ml of cold ethyl acetate and chromatographed as described by Gorman et al. (1982a). For quantitation of CAT activity, reactions were in the linear range (no more than 40% Cm acetylated).

(d) RNA isolation and analysis

Total RNA was extracted from pooled injected oocytes and digested with proteinase K primarily as described (Probst et al., 1979; Colman, 1984a). Homogenate was extracted twice with phenol–chloroform–isoamyl alcohol (20 : 20 : 1, v/v) and once with chloroform–isoamyl alcohol. After precipitation with ethanol, the RNA was collected by centrifugation and resuspended in 10 mM Tris · HCl (pH 7.4), 0.1 mM EDTA, 0.1% SDS.

Primer extension analyses were performed as previously described (Hu et al., 1986) to detect accumulated *tk* transcripts in total RNA purified from injected oocytes. The 33-nt ^{32}P-labeled synthetic oligodeoxynucleotide 5′-dGTGCGGGAGTTT-CACGCCACCAAGATCTGCGGC-3′ (synthesized by S. Horvath, Caltech, on an automated DNA synthesizer) was used as primer to hybridize to RNA transcribed from the pTK-5 template. The synthetic oligodeoxynucleotide is complementary to a region of the *tk* mRNA started at 44 nt downstream from the *tk* mRNA cap site. Therefore, the primer extension product observed from RNA transcribed from the pTK-5 template is 77 nt in length. Extension products were sized by electrophoresis on a 9% PA–8 M urea sequencing gel in Tris-borate-EDTA buffer, and detected by autoradiography of the dried gel. Primer extension mapping of RNA transcribed from the MSV-*cat* template was performed as previously described (Hu and Davidson, 1987).

RESULTS

(a) The *lac* operator-repressor complex functions as a regulatable switch in *Xenopus* oocytes

The transcriptional units used in these experiments (Fig. 1A) were also used by us in our recent studies of the *lac* operator-repressor system in mammalian cells (Hu and Davidson, 1987). In these constructions, the *lac* operator sequence has been inserted into a selected site in the region of the promoter or tsp of a plasmid, pSM11. In pSM11, the *E. coli cat* gene coding sequence is transcribed under the control of a chimeric promoter consisting of part of the SV40 early promoter, coupled to the MSV enhancer (Laimins et al., 1984). SV40 RNA processing signals, the small-t intron and early poly(A)-addition signal, are provided at the 3′ end of the gene. This particular construction was useful since the effect of the MSV enhancer in augmenting expression in mammalian cells had already been reported (Laimins et al., 1984).

As can be seen in Fig. 1A, in the plasmid pSMEO1 an operator sequence has been inserted 16 nt downstream from the second A of the TATA box and therefore 9 nt upstream from the tsp. In the plasmid pSMDO1, the operator is inserted 40 nt downstream from the tsp, thus lying 32 nt upstream

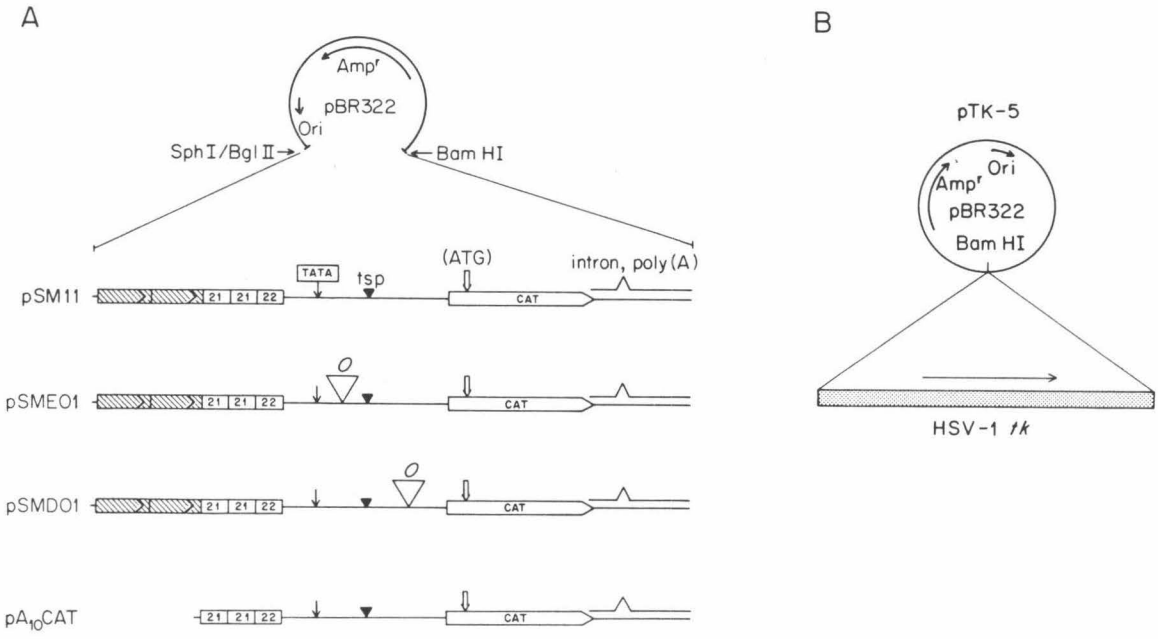


Fig. 1. Schematic representation of structure of operator insertions in the promoter regions and plasmid pTK5. (**A**) The organization of the SV40 early promoter region in parental plasmid (pSM11) and in its derivatives containing the *lac* operator insertions is depicted. The hatched boxes to the left of the SV40 promoter (21, 21, 22) illustrate the MSV enhancer (the 73/72 bp tandem repeats; Laimins et al., 1984). The SV40 21-bp repeats, TATA box, tsp, start codon (ATG), CAT-coding sequences (boxed region), the SV40 small-t intron (intron) and early poly (A)-addition signal are indicated. The positions of the *lac* operator (*O*) insertion are shown as triangular inserts. Plasmid pSMEO1 carries one operator insertion between the TATA box region and the tsp, whereas plasmid pSMDO1 carries one operator insertion between the tsp and the start codon (ATG). As a control for the activity of the MSV enhancer in *Xenopus* oocytes, the plasmid $pA_{10}CAT$ that contains the SV40 promoter elements but no enhancer sequences is used. (**B**) Structure of plasmid pTK-5. The 3.5-kb *Bam*HI fragment containing the HSV-1 *tk* gene (stippled box) is inserted into the *Bam*HI site of pBR322 vector. Amp^r, ampicillin-resistance (*bla*) gene. Ori, origin of DNA replication.

from the A of the *cat* gene initiator (ATG). We first tested whether CAT enzyme activity could be detected in *Xenopus* oocytes after injection of these plasmids into the nucleus under the usual conditions (5 ng of each DNA injected per oocyte) (Gurdon and Melton, 1981; Gurdon and Wickens, 1983). As a control (see below), the plasmid pTK-5 (Fig. 1B) containing a HSV-1 *tk* gene was coinjected with the pSM plasmids. As shown in Fig. 2A, Exp. 1 (each set of two lanes shows duplicate experiments) and in Fig. 3A, Exp. 1, extracts prepared from pSMEO1-injected oocytes and from pSMDO1-injected oocytes, respectively, contained readily detected amounts of CAT enzyme. When the same mixture of either plasmid was premixed with the purified *E. coli* *lac* repressor protein (at a ten-fold molar excess over the pSM plasmid), and injected into the oocyte nucleus, *cat* gene expression was undetectable (Fig. 2A, Exp. 2 [the second set of two lanes]; and Fig. 3A, Exp. 2). No CAT enzyme activity was detected in uninjected oocytes, confirming the observation by Jones et al. (1983) that *Xenopus* oocytes are devoid of endogenous CAT activity. It should be noted that oocytes from different frogs have a high degree of variability in their ability to synthesize the CAT enzyme (Jones et al., 1983). Thus, the CAT activities from these two experiments (pSMEO1 and pSMDO1) are not directly comparable.

In *E. coli*, IPTG added to the external medium is an efficient inducer of *lac* gene expression from the repressed state, we therefore tested whether IPTG would also relieve repression in oocytes. As shown in Fig. 2A, Exp. 4 (the fourth set of two lanes), and Fig. 3A, Exp. 4, injection of DNA plus repressor into oocyte nuclei, followed by incubation in an external medium containing 20 mM IPTG for about

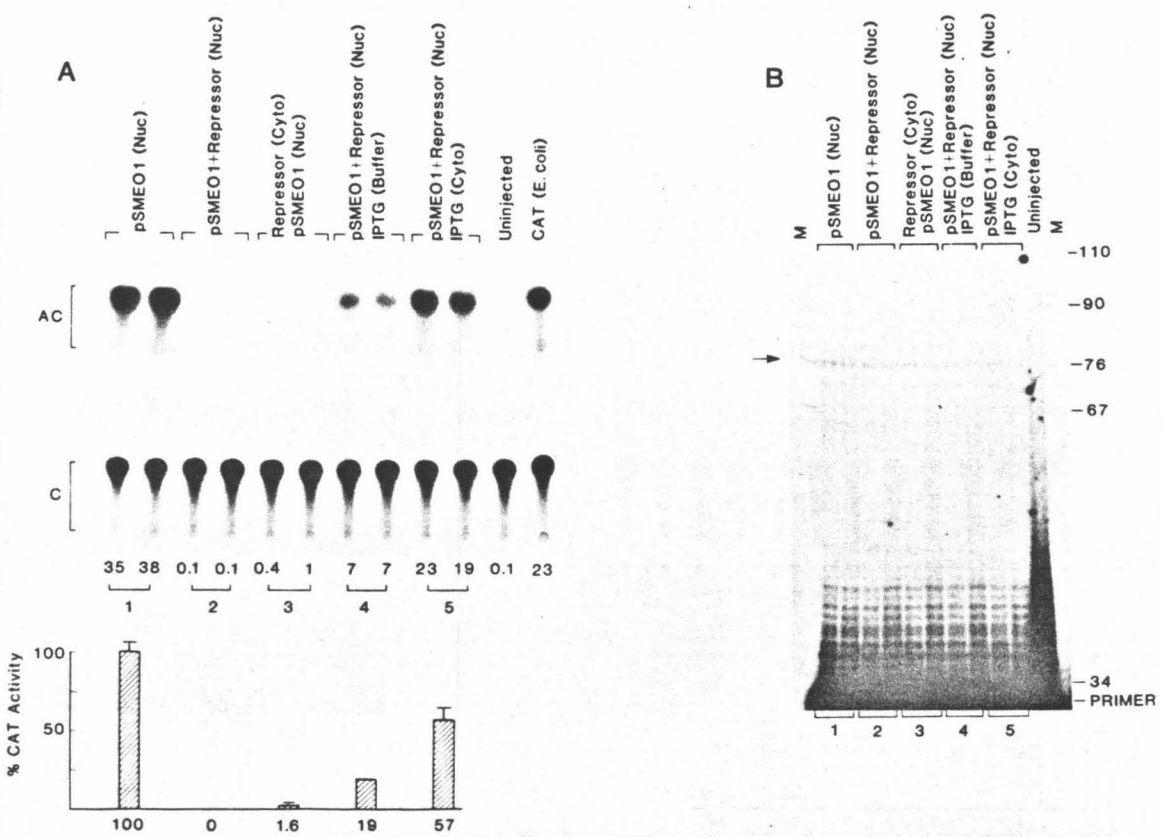


Fig. 2. Effect of *lac* repressor on the *cat* expression from plasmid pSMEO1 and mapping of transcripts. (**A**) CAT synthesis in the absence or presence of the *lac* repressor in injected oocytes. Groups of 40 oocytes isolated from a single female frog were each coinjected with plasmids pSMEO1 and pTK-5 (5 ng of each DNA per oocyte). After 48 h of incubation, total extracts were isolated from two batches of pooled oocytes (ten oocytes per batch) and the extracts from each batch equivalent to one oocyte were assayed for CAT activity (experiments 1-5). Specifically, in Exp. 1, plasmids pSMEO1 and pTK-5 were injected into the oocyte nucleus. In Exp. 2 plasmids pSMEO1 and pTK-5 were coinjected with the *E. coli lac* repressor (at a ten-fold molar excess over the pSM plasmid) into the oocyte nucleus. In Exp. 3, the *lac* repressor (at approx. 70 ng protein per oocyte) was first injected into the oocyte cytoplasm and, 2 h later, plasmids pSMEO1 and pTK-5 were injected into the nucleus of the previously injected oocytes. In Exp. 4, oocytes were injected as described in Exp. 2 and subsequently incubated in the buffer containing 20 mM IPTG. In Exp. 5, oocytes were first injected as described in Exp. 2 and, 12 h later, IPTG was injected into the cytoplasm (to a final concentration about 20 mM in oocyte) of previously injected oocytes. CAT assays were performed under linear assay conditions. CAT assays with extracts from uninjected oocytes and with the purified *E. coli* CAT enzyme were performed as negative and positive controls, respectively. The figure shows an autoradiograph of thin-layer chromatography plates used to analyze the products of the reaction. The migration position of the starting material, Cm is indicated (row C). The migration position of acetylated Cm is indicated (row AC). The number below each lane indicates the % conversion of Cm to its acetylated forms, as quantified by liquid scintillation counting. The diagram shown in the lower panel is the relative CAT activity which is normalized as percentage of plasmid pSMEO1 CAT activity in injected oocytes in Exp. 1. The background radioactivity of CAT assays, as shown in the negative control, has been subtracted. (**B**) Primer-extension mapping of transcripts synthesized in injected oocytes by the HSV-1 *tk* gene. All injections are described above. After 48 h of incubation, total RNAs were isolated from two batches of pooled oocytes (10 oocytes per batch) and the RNAs equivalent to one oocyte were analyzed for *tk* transcripts by primer extension. The HSV-1 *tk* gene produced a transcript that resulted in the 33-nt *tk* specific primer being extended to a length of 77 nt (see MATERIALS AND METHODS, section **d**). Extension products were sized on a 9% PA–8 M urea sequencing gel and autoradiographed. As a control, the same primer extension experiment was performed by using total RNA isolated from pooled uninjected oocytes. ^{32}P-labeled *Hpa*II fragments of pBR322 served as nt size markers (M), which are marked in the right margin. The bands corresponding to the *tk* transcript-specific products are indicated by an arrow on the left margin. Nuc, nucleus; Cyto, cytoplasm; Buffer, ND-96 buffer containing 20 mM IPTG; Uninjected, uninjected oocytes; PRIMER, the 33-nt *tk*-specific primer.

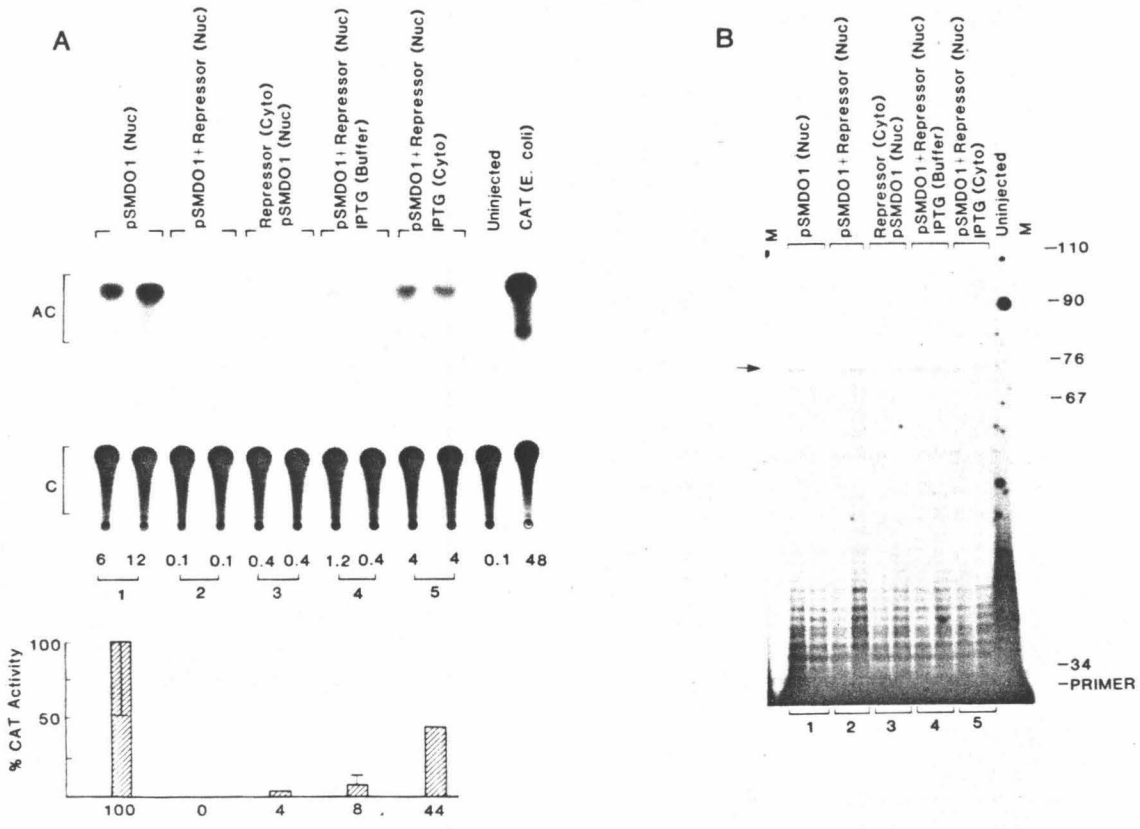


Fig. 3. Efffect of *lac* repressor on the *cat* expression from plasmid pSMDO1 and mapping of transcripts. (**A**) CAT synthesis in the absence or presence of the *lac* repressor in injected oocytes. Groups of 40 oocytes isolated from another single female frog were each coinjected with plasmids pSMDO1 and pTK-5 (5 ng of each DNA per oocyte). Experimental procedures and CAT assays with extracts isolated from two batches of pooled oocytes of each experiment were performed as described in the legend to Fig. 2A. Specifically, in Exp. 1, plasmids pSMDO1 and pTK-5 were injected into the oocyte nucleus. In Exp. 2, plasmids pSMDO1 and pTK-5 were coinjected with the purified *lac* repressor (at a ten-fold molar excess over the pSM plasmid) into the oocyte nucleus. In Exp. 3, the *lac* repressor was first injected into the oocyte cytoplasm (at approx. 70 ng per oocyte) and, 2 h later, plasmids pSMDO1 and pTK-5 were injected into the previously injected oocyte nucleus. In Exp. 4, oocytes were injected as described in Exp. 2 and subsequently incubated in the buffer containing 20 mM IPTG. In Exp. 5, oocytes were first injected as described in the Exp. 2 and, 12 h later, IPTG was injected into the cytoplasm of the previously injected oocytes (to a final concentration of approx. 20 mM in oocyte). CAT assays were performed under linear assay conditions. All symbols and the diagram of the relative CAT activity shown in the lower panel are the same as in Fig. 2A. (**B**) Primer extension mapping of transcripts synthesized in injected oocytes by the HSV-1 *tk* gene. All injections are described above. Experimental procedures and primer extension analyses for the *tk*-specific transcripts were performed as described in the legend to Fig. 2B. Symbols are the same as in Fig. 2B.

48 h, induced expression of CAT activity of pSMEO1 to approx. 19% and CAT activity of pSMDO1 to approx. 8% of the unrepressed level. Thus, some relief of repression was achieved. We suspected that this limited induction was due to limited permeability of the oocyte membrane to the inducer IPTG. Therefore, 12 h after the pSM plasmid DNAs plus repressor were injected into the nucleus, IPTG was injected into the cytoplasm (to a final average concentration of 20 mM in the oocyte). This protocol was chosen to ensure that any IPTG induction was due to an effect on operator-bound repressor, rather than inhibition of the initial binding of repressor to operator. As shown in Fig. 2A, Exp. 5

(the fifth set of two lanes), and Fig. 3A, Exp. 5, injection of IPTG into the cytoplasm induced expression of *cat* gene by pSMEO1 up to 57%, and by pSMDO1 up to 44%, of the unrepressed level. To increase the induction level to 100%, we have tried to inject more IPTG into the cytoplasm (up to a final concentration of 50 mM), but these concentrations had an adverse effect on the health of the oocyte. Thus, while the present results show that IPTG can significantly relieve repression, we have not yet been able to induce complete dissociation of operator-bound repressor molecules by IPTG. IPTG does not have non-specific effects on the expression of the HSV-1 *tk* gene (Figs. 2B and 3B).

(b) Functional *lac* repressor can enter the oocyte nucleus

Athough the *lac* repressor is a well-characterized DNA-binding and regulatory protein in *E. coli*, it was not known whether it could penetrate the oocyte nuclear envelope and bind to any operator containing sequences therein. To address this question, we injected the *lac* repressor into the cytoplasm of oocytes (at approx. 70 ng protein per oocyte), and then 2 h later, injected plasmids pSMEO1 and pTK-5 (5 ng of each DNA per oocyte) into the nucleus of these cells. As shown in Fig. 2A, Exp. 3 (the third set of two lanes), and Fig. 3A, Exp. 3, injection of repressor proteins into the oocyte cytoplasm prior to injection of DNAs into the oocyte nucleus reduced expression of CAT activity of pSMEO1 by 98% and CAT activity of pSMDO1 by 96%. The present experiments therefore show that a sufficient quantity of repressor can enter the nucleus and bind to the operator sequence to give nearly complete repression.

(c) The nonspecific DNA-binding properties of the *lac* repressor do not contribute to transcriptional inhibition

It is known from previous studies (Lin and Riggs, 1972; 1975; von Hippel et al., 1974) that the *lac* repressor can bind nonspecifically to DNA. Therefore it was important to determine whether any such nonspecific binding would affect gene expression in our system. These tests were conducted by coinjecting the HSV-1 *tk* gene into the oocyte nucleus along with the pSM plasmids in the absence, and presence, of repressor protein. After 48 h of incubation, the amount of *tk*-specific RNA molecules were measured by primer extension (see legend to Fig. 2B). As shown in Figs. 2B and 3B (each set of two lanes showing duplicate experiments), there was no significant difference in the amount of *tk*-transcripts synthesized between oocytes injected with and without the *lac* repressor. Thus, the non-specific DNA-binding properties of the repressor do not have an inhibitory effect on transcription of a gene introduced by injection.

In separate injections, we also tested that expression from the parental plasmid, pSM11, that contains no *lac* operator insert, was not inhibited by *lac* repressor (not shown). Because of the variability of oocyte injection experiments, we prefer to use internal controls, i.e., the HSV-1 *tk* gene (Figs. 2B and 3B).

(d) Effects of operator insertions on gene expression in *Xenopus* oocytes

As a further control, it was important to determine whether insertion of the *lac* operator sequence up-

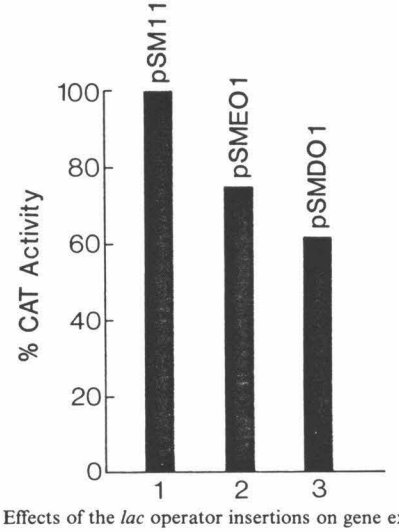


Fig. 4. Effects of the *lac* operator insertions on gene expression in the absence of repressor in oocytes. Groups of 20 oocytes isolated from a single female frog were each injected (5 ng of DNA per oocyte) with one of these plasmids pSM11 (lane 1), pSMEO1 (lane 2), and pSMDO1 (lane 3). CAT assays with extracts isolated from pooled injected oocytes of each experiment were performed as described in the legend to Fig. 2A. The relative CAT activity is normalized as % of parental plasmid pSM11 CAT activity in injected oocytes.

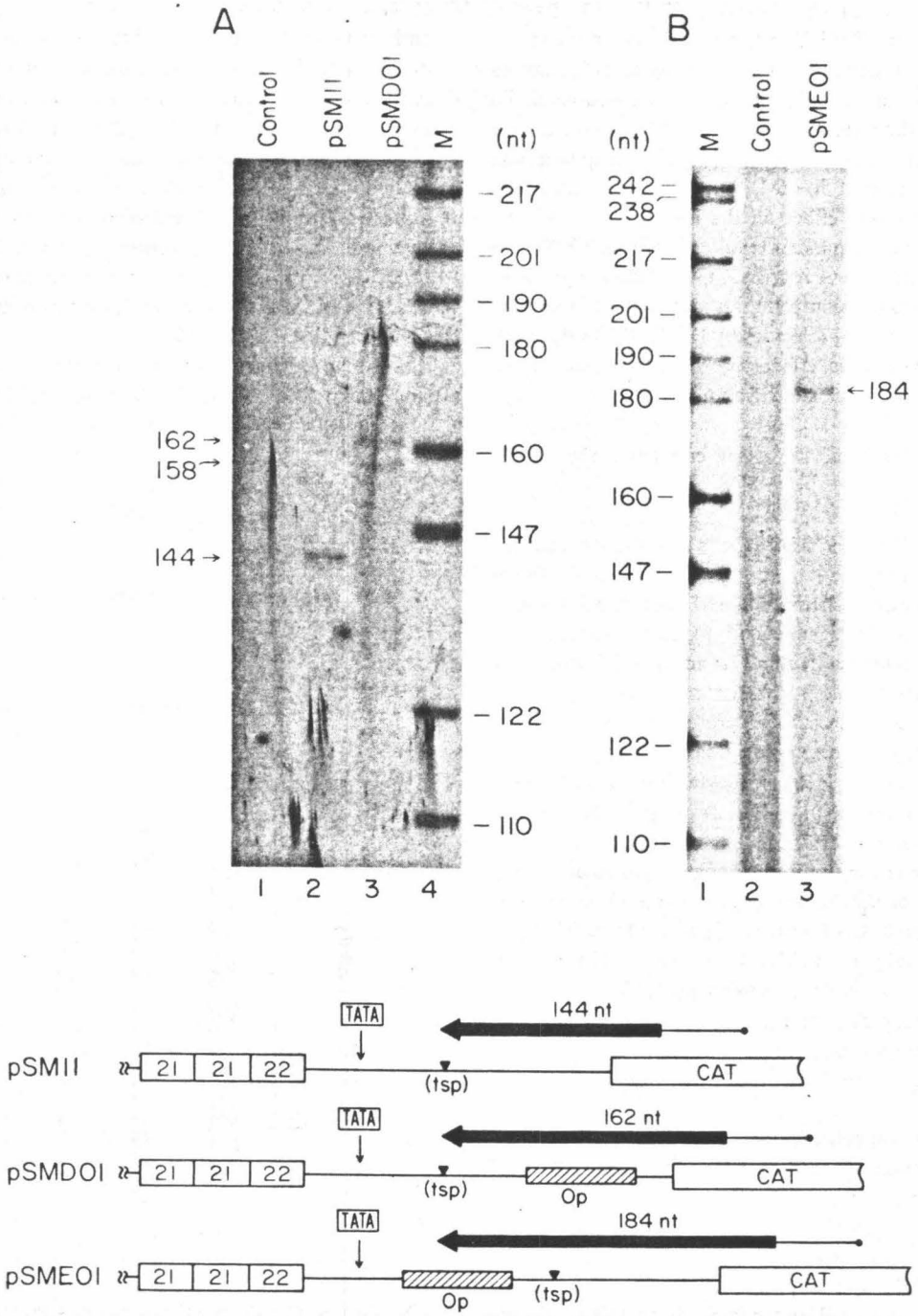


Fig. 5. Primer extension mapping of transcripts synthesized in injected oocytes from MSV-*cat* fusion genes containing operator insertions. After 48 h of incubation, total RNAs were isolated from ten pooled injected oocytes, and the RNAs equivalent to five oocytes were analyzed for *cat* gene transcripts by primer extension. The 5′-^{32}P-labeled 36-nt synthetic oligodeoxynucleotide complementary to the CAT-coding sequence (Hu and Davidson, 1987) was hybridized to the RNA and then extended with unlabeled dNTPs with reverse

stream from the *cat* gene affects transcription in *Xenopus* oocytes in the absence of repressor. Such an analysis was provided by comparison of expression in the absence of repressor by the plasmids pSMEO1 and pSMDO1, which contain an operator insert, with the parental plasmid pSM11 which contains no operator insert. As shown in Fig. 4, the plasmids pSMEO1 and pSMDO1 show a reduction of approx. 25% and 38%, respectively, relative to CAT activity expression by the plasmid pSM11, which contains no operator sequence. Thus, the operator sequence itself has a modest reducing effect on the levels of *cat* gene expression in *Xenopus* oocytes. This is similar to our previous results in mammalian cells (Hu and Davidson, 1987).

(e) Operator insert immediately downstream of the TATA box shifts the transcription start point of an injected gene in *Xenopus* oocytes

Whenever one inserts a new DNA segment close to the TATA box or the tsp, and then observes changes in gene expression, it is important to show whether or not the tsp has changed. It was, therefore, desirable to determine whether the operator insertions in the plasmids studied here would influence the choice of the tsp in *Xenopus* oocytes.

To test this, we performed *cat*-specific primer extension mapping with total RNA extracted from injected oocytes with each MSV-*cat* fusion gene tested. We used as primer a 36-nt 5′-end-labeled oligodeoxynucleotide complementary to a sequence within the *cat* gene, as previously described (Hu and Davidson, 1987). As shown in Fig. 5A (lane 2), with RNA from oocytes injected with the parental plasmid pSM11, the DNA primer was extended to a length of 144 nt. However, when RNA from oocytes injected with the plasmid pSMEO1, containing one operator segment of length 40-bp insertion between the TATA box and the tsp, the same primer was extended to 184 nt (Fig. 5B, lane 3). This corresponds to a new tsp at a T residue, 25 bp downstream from the second A of the TATA box, within the operator sequence. In the plasmid pSMDO1, an operator segment of length 40 bp had been substituted for a 21-bp fragment between the tsp and the start codon ATG (see MATERIALS AND METHODS, section **a**). With RNA from oocytes injected with the plasmid pSMDO1, the primer extension products were extended to approx. 162 and 158 nt (Fig. 5A, lane 3). The latter extension product (158 nt) may be due to a small secondary structure in the 5′-untranslated region of the mRNA (or it may be a nearby alternate tsp), but the fragment of length 162 nt corresponds to the same tsp as for the parental plasmid. We therefore conclude that an operator insertion downstream from the normal tsp does not affect the position of the actual tsp in oocytes. It should be noted that the relative intensities of these extended products are not comparable, since these injected oocytes were not from the same frog.

(f) The MSV enhancer has an activating effect on an enhancerless SV40 early promoter expression in *Xenopus* oocytes

The MSV enhancer was first discovered by its ability to functionally replace the SV40 enhancer (72-bp repeats) during viral infection (Levinson et al., 1982). Later it was demonstrated that, although the MSV enhancer is capable of enhancing gene activity, both from its own promoter (Laimins et al., 1984; Graves et al., 1985) and from a heterologous (SV40) promoter in mammalian cells (Laimins et al., 1982; 1984), it does not contribute to transcriptional expression from its own promoter in *Xenopus* oocytes (Graves et al., 1985). The failure of the MSV enhancer to enhance its own promoter in *Xenopus* oocytes was interpreted to reflect a species and cell-type specificity limit of enhancer function (for reviews, see Gluzman, 1985; and Serfling et al., 1985). However, it has been demonstrated that the

transcriptase. Extended products were sized on a 6% PA sequencing gel (Maxam and Gilbert, 1980) and autoradiographed. As a control, the same primer was hybridized to the total RNA isolated from pooled uninjected oocytes. (**A**) Lanes: 1, control: RNA from uninjected oocytes; 2, RNA from oocytes injected with pSM11; 3, RNA from oocytes injected with pSMDO1; 4, ^{32}P-labeled *Hpa*II fragments of pBR322 served as nt size markers (M), which are inidicated. (**B**). Lanes: 1, nt size markers; 2, control: RNA from uninjected oocytes; 3, RNA from oocytes injected with pSMEO1. The band due to the extended product is identified by an arrow, and its length is indicated. Synthesis of the observed cDNA products is depicted schematically below the figure. Symbols are the same as in the legend to Fig. 1, except that the *lac* operator insertion (Op) is indicated by a hatched box. The primer is shown as a thin line, and extension products as thick horizontal arrows.

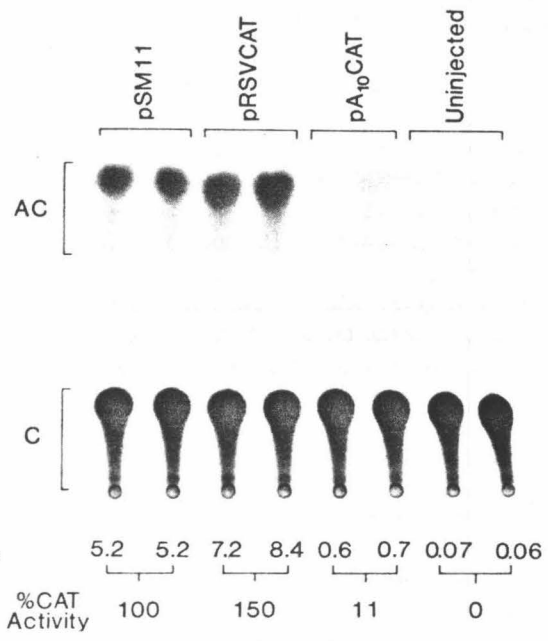


Fig. 6. Activating effect of the MSV enhancer on an enhancer-less SV40 early-promoter expression and strong expression of the RSV LTR in injected oocytes. Groups of 40 oocytes isolated from a single female frog were each injected with plasmid DNA (5 ng DNA per oocyte) as indicated in each lane. After 48 h incubation, total extracts were isolated from two batches of pooled injected oocytes (ten oocytes per batch) and the extracts from each batch equivalent to one oocyte were assayed for CAT activity. CAT assays were performed under linear assay conditions. CAT assays with extracts from uninjected oocytes were performed as negative controls. Symbols are as described in the legend to Fig. 2A. The number below each lane indicates the % conversion of Cm to its acetylated forms, as quantified by liquid scintillation counting. The number shown in the bottom row is the relative CAT activity which is normalized as % of plasmid pSM11 CAT activity in injected oocytes. The background radioactivity of CAT assays (as shown in negative controls: lanes marked uninjected) has been subtracted.

SV40 enhancer is able to activate transcription about 100-fold from the SV40 early promoter in *Xenopus* oocytes (Spinelli and Ciliberto, 1985). Since the SV40 enhancer is known to be active in a wide variety of tissues and species, it was of interest to examine whether or not the MSV enhancer can functionally replace the SV40 enhancer to activate transcription from the SV40 early promoter in *Xenopus* oocytes. We therefore compared *cat* gene expression in injected oocytes from the MSV-*cat* fusion plasmid pSM11 with that from the plasmid $pA_{10}CAT$, containing the SV40 early promoter elements but no enhancer. To test whether the RSV LTR is an efficient promoter for gene expression in frog oocytes, we also performed CAT assays with extracts from oocytes injected with the plasmid pRSVCAT, in which the *cat* gene is placed under RSV LTR control (Gorman et al., 1982b). CAT assays with extracts from uninjected oocytes were performed as a negative control. As shown in Fig. 6 (with extracts from pooled oocytes, and again with each set of two lanes showing duplicate experiments), the level of *cat* gene expression from plasmid pSM11 was nine times higher than that from plasmid $pA_{10}CAT$, and CAT activity of pRSVCAT was about 14 times higher than that of $pA_{10}CAT$. Therefore, we conclude that the MSV enhancer has a significant activating effect on SV40 early promoter expression, and that the RSV LTR is indeed a strong promoter for gene expression in *Xenopus* oocytes.

DISCUSSION

We have shown that with a single *lac* operator sequence inserted in a control region of the test plasmid injected into *Xenopus* oocytes, a high level of repression is achieved with the *lac* repressor. Repression can be achieved by premixing repressor with plasmid prior to nuclear injection, or by injection of repressor into the oocyte cytoplasm before injection of the plasmid into the oocyte nucleus. A low level of induction can be achieved by addition of the inducer IPTG to the external medium; a much higher level of induction is obtained by injection of IPTG into the oocyte cytoplasm. The repressor protein does not affect expression of a test plasmid that does not contain the operator sequence. These points are discussed in greater detail below.

A primary goal of this study was to demonstrate that the inducible *lac* operator-repressor system could be used as a genetic switch in frog oocytes, just as it is in *E. coli* and in mammalian cells (Hu and Davidson, 1987; Brown et al., 1987). The ability to control the transcription of a gene in oocytes by using the *lac* repressor and hybrid promoters containing the *lac* operator should be useful for a wide variety

of purposes. For example, it is often desirable to regulate the expression of a protein in an inducible manner rather than rely on constitutive expression since the expression of high levels of proteins may have deleterious effects on the growth of the cells (Kaufman et al., 1985). Furthermore, one potentially rewarding feature of this system is that it may provide a method for the experimenter to control the expression of genes that play a role in early embryonic development.

Our previous studies and those of Brown et al. (1987) showed that in mammalian cells, the *lac* repressor synthesized in the cytoplasm could enter the nucleus; the concentration of repressor was approximately equal between nucleus and cytoplasm (Hu and Davidson, 1987). The present results show that *lac* repressor protein injected into the oocyte cytoplasm can migrate into the nucleus and cause repression. In this instance, it was impracticable to determine the distribution of repressor between the cytoplasm and the nucleus, thus our results do not indicate whether or not specific binding increases the repressor concentration in a specific nuclear or cellular compartment.

A significant and potentially useful difference between the oocyte experiments and those in mammalian cells (Hu and Davidson, 1987) is the higher level of repression achieved in oocytes. This is probably due to a higher concentration of repressor protein in injected oocytes. Since the volume of a *Xenopus* oocyte and a mammalian cell are approximately 0.54 μl (Gurdon and Wickens, 1983) and 4×10^{-6} μl (Alberts et al., 1983), respectively, we estimate that we have introduced about an 83 times higher concentration of repressor in an oocyte (approx. 830 nM) than that in a repressor-producing LI-1 cell (approx. 10 nM) (the repressor-expressing L cells studied in Hu and Davidson, 1987). The average concentration of plasmid DNA in an injected oocyte nucleus (approx. 30 nM) is about 30 times higher than that transfected into a nucleus of the LI-1 cell (approx. 1 nM). Thus, the molar ratio of repressor versus plasmid DNA in injected oocytes is approx. three times higher than that in our transient expression gene transfer experiments in LI-1 cells. This would suggest that higher levels of repression may be achieved by increasing the total concentration of repressor in an animal cell over that in the LI-1 cell line.

We wish to discuss the interpretation of the observation that the *lac* repressor does not affect transcription from a gene that does not contain an operator sequence. This fact may be due to the relative weakness of the nonspecific binding, so that RNA polymerase can displace nonspecifically bound repressor or it may indicate that the *lac* repressor simply does not bind nonspecifically to DNA in nucleosomes. The latter interpretation is supported by the observations of Chao et al. (1980). They have shown that assembly of *lac* operator-containing DNA into nucleosomes with octameric histone cores in vitro does not affect specific sequence recognition by repressor, or IPTG-induced dissociation, but that nonspecific binding of repressor to DNA is considerably diminished. On the other hand, the specific repression results show that the repressor can bind to operator DNA in an oocyte nucleus. We think it is possible that this result means that for an operator segment in a nucleosome, the major groove of the operator sequence faces outward and is available for recognition by repressor, although the minor groove is interacting with the histone core. The alternative possibility that operator sequences are not organized in nucleosomes in oocytes has not been investigated by us.

In agreement with our earlier results with mammalian cells, binding of repressor to an operator inserted downstream from the tsp inhibits transcription. This is probably not due to a direct inhibition of binding of RNA polymerase (see below), but rather, to a block of the propagating transcription complex. The latter interpretation is supported by observations in *E. coli* that the operator-bound repressor can block transcribing RNA polymerase and terminate transcription (Deuschle et al., 1986; Besse et al., 1986; Sellitti et al., 1987). In the *E. coli lac* operon, the operator overlaps downstream with the promoter element. When repressor is bound to the operator, RNA polymerase can still bind to the promoter but can not initiate transcription (Schmitz and Galas, 1979). In the expression plasmid we have studied here, pSMDO1, the operator is inserted at 40 bp downstream from the normal tsp (65 bp downstream from the TATA box). By occupying this downstream operator, the repressor is less likely to provide any steric hindrance and to prevent the promoter from being recognized by RNA polymerase. However, we do not have data to show where an

operator-bound repressor forces the transcribing RNA polymerase to stop and to release the nascent transcript. From the standpoint of basic transcription studies, one exciting feature of this system is that it provides for the potential study of how various distant control elements, such as enhancers, communicate with the promoter and the tsp of a eukaryotic gene.

We wish to discuss an incidental but important set of observations about the activity of the MSV enhancer that were made in the course of this work. Previously, it has been reported that the MSV enhancer does not appear to be recognized or utilized to enhance transcription from its own promoter in *Xenopus* oocytes (Graves et al., 1985). In contrast, our results show that the MSV enhancer placed upstream from an enhancerless SV40 early promoter does activate transcription in *Xenopus* oocytes. Thus, *Xenopus* oocytes do contain a *trans*-acting factor(s) that can act on the MSV enhancer to enhance expression from some *cis*-located promoters. Interestingly, it has been shown recently that in human 293 cells the SV40 enhancer has only a minor effect on transcription from the SV40 early promoter, but that it has a strong activating effect on transcription from the HSV *tk* promoter (Robbins et al., 1986). Overall these results, taken together, support the view that the interactions between an enhancer, a promoter, and other *cis*-acting DNA elements will vary with the particular *trans*-acting factor(s) that are present in the cell of interest, but will also depend on the particular permutation of *cis*-acting DNA elements present in the total promoter cassette.

ACKNOWLEDGEMENTS

We are grateful to the late Dr. George Khoury (National Cancer Institute) for generously providing plasmid pSM11, to Drs. Kathleen Matthews (Rice University) and Sheldon York (University of Denver) for gifts of the purified *E. coli lac* repressor. M.H. thanks Dr. Terry Snutch (Caltech) for his excellent instruction on *Xenopus* oocyte microinjection methods, providing some oocytes, and helpful comments on this work and the manuscript; Frank Preugschat (Caltech) for his instruction on oocyte nuclear injections, and Dr. Henry Lester (Caltech) for the use of the microinjection equipment. This work was supported by a research grant from the National Institutes of Health.

REFERENCES

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. and Watson, J.D.: Molecular Biology of the Cell. Garland, New York, 1983, p. 92.

Alwine, J.C.: Transient gene expression control: effects of transfected DNA stability and *trans*-activation by viral early proteins. Mol. Cell. Biol. 5 (1985) 1034–1042.

Beckwith, J.R. and Zipser, D.: The Lactose Operon. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1970.

Besse, M., von Wilcken-Bergmann, B. and Müller-Hill, B.: Synthetic *lac* operator mediates repression through *lac* repressor when introduced upstream and downstream from *lac* promoter. EMBO J. 5 (1986) 1377–1381.

Brown, M., Figge, J., Hansen, U., Wright, C., Jeang, K., Khoury, G., Livingston, D.M. and Roberts, T.M.: *lac* repressor can regulate expression from a hybrid SV40 early promoter containing a *lac* operator in animal cells. Cell 49 (1987) 603–612.

Chao, M.V., Gralla, J.D. and Martinson, H.G.: *lac* operator nucleosomes, I. Repressor binds specifically to operator within the nucleosome core. Biochemistry 19 (1980) 3254–3260.

Colman, A.: Expression of exogenous DNA in *Xenopus* oocytes. In Hames, B.D. and Higgins, S.J. (Eds.), Transcription and Translation: A Practical Approach. IRL Press, Oxford, 1984a, pp. 49–69.

Colman, A.: Translation of eukaryotic messenger RNA in *Xenopus* oocytes. In Hames, B.D. and Higgins, S.J. (Eds.), Transcription and Translation: A Practical Approach. IRL Press, Oxford, 1984b, pp. 271–302.

Deuschle, U., Gentz, R. and Bujard, H.: *lac* repressor blocks transcribing RNA polymerase and terminates transcription. Proc. Natl. Acad. Sci. USA 83 (1986) 4134–4137.

Domen, J., Van Leen, R.W., Lubsen, N.H. and Schoenmakers, J.G.G.: A vital staining method for measuring the efficiency of transfection of eukaryotic cells. Anal. Biochem. 155 (1986) 379–384.

Dumont, J.: Oogenesis in *Xenopus laevis* stages of oocyte development in laboratory maintained animals. J. Morphol. 136 (1972) 153–180.

Gluzman, Y.: Eukaryotic Transcription. The Role of *cis*- and *trans*-acting Elements in Initiation. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1985.

Gorman, C.M., Moffat, L.F. and Howard, B.H.: Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. Mol. Cell. Biol. 2 (1982a) 1044–1051.

Gorman, C.M., Merlino, G.T., Willingham, M.C., Pastan, I. and Howard, B.H.: The Rous sarcoma virus long terminal repeat is a strong promoter when introduced into a variety of eukaryotic cells by DNA-mediated transfection. Proc. Natl. Acad. Sci. USA 79 (1982b) 6777–6781.

Graves, B.J., Eisenman, R.N. and McKnight, S.L.: Delineation of transcriptional control signals within the Moloney murine sarcoma virus long terminal repeat. Mol. Cell. Biol. 5 (1985) 1948–1958.

Grosschedl, R. and Birnstiel, M.L.: Identification of regulatory sequences in the prelude sequences of an H2A histone gene by the study of specific deletion mutants in vivo. Proc. Natl. Acad. Sci. USA 77 (1980) 1432–1436.

Gurdon, J.B. and Melton, D.A.: Gene transfer in amphibian eggs and oocytes. Annu. Rev. Genet. 15 (1981) 189–218.

Gurdon, J.B. and Wickens, M.P.: The use of *Xenopus* oocytes for the expression of cloned genes. Methods Enzymol. 101 (1983) 370–386.

Hu, M.C.-T. and Davidson, N.: The inducible *lac* operator-repressor system is functional in mammalian cells. Cell 48 (1987) 555–566.

Hu, M.C.-T., Sharp, S.B. and Davidson, N.: The complete sequence of the mouse skeletal α-actin gene reveals several conserved and inverted repeat sequences outside of the protein-coding region. Mol. Cell. Biol. 6 (1986) 15–25.

Jones, N.C., Richter, J.D., Weeks, D.L. and Smith, L.D.: Regulation of adenovirus transcription by an E1a gene in microinjected *Xenopus laevis* oocytes. Mol. Cell. Biol. 3 (1983) 2131–2142.

Kaufman, R.J., Wasley, L.C., Spiliotes, A.J., Gossels, S.D., Latt, S.A., Larson, G.R. and Kay, R.M.: Coamplification and coexpression of human tissue-type plasminogen activator and murine dihydrofolate reductase sequences in Chinese hamster ovary cells. Mol. Cell. Biol. 5 (1985) 1750–1759.

Laimins, L.A., Khoury, G., Gorman, C., Howard, B. and Gruss, P.: Host-specific activation of transcription by tandem repeats from simian virus 40 and Moloney murine sarcoma virus. Proc. Natl. Acad. Sci. USA 79 (1982) 6453–6457.

Laimins, L.A., Gruss, P., Pozzatti, R. and Khoury, G.: Characterization of enhancer elements in the long terminal repeat of Moloney murine sarcoma virus J. Virol. 49 (1984) 183–189.

Levinson, B., Khoury, G., Vande Woude, G. and Gruss, P.: Activation of SV40 genome by 72-base pair tandem repeats of Moloney sarcoma virus. Nature 295 (1982) 568–572.

Lin, S. and Riggs, A.D.: *lac* repressor binding to non-operator DNA: detailed studies and a comparison of equilibrium and rate competition methods. J. Mol. Biol. 72 (1972) 671–690.

Lin, S. and Riggs, A.D.: The general affinity of *lac* repressor for *E. coli* DNA: implications for gene regulation in procaryotes and eucaryotes. Cell 4 (1975) 107–111.

Maxam, A.M. and Gilbert, W.: Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. 65 (1980) 499–559.

McKnight, S.L. and Gavis, E.R.: Expression of the herpes thymidine kinase gene in *Xenopus laevis* oocytes: an assay for the study of deletion mutants constructed in vitro. Nucl. Acids Res. 8 (1980) 5931–5948.

McKnight, S. and Kingsbury, R.: Transcriptional control signals of a eukaryotic protein-coding gene. Science 217 (1982) 316–324.

Miller, J.H. and Reznikoff, W.S.: The Operon. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1980.

Probst, E., Kressmann, A. and Birnstiel, M.L.: Expression of sea urchin histone genes in the oocyte of *Xenopus laevis*. J. Mol. Biol. 135 (1979) 709–732.

Roach, A., Nicolson, M.O. and Davidson, N.: Transcription and expression of the herpes simplex virus *tk* gene inserted into proviral sequences of feline leukemia virus. Gene 32 (1984) 389–398.

Robbins, P.D., Rio, D.C. and Botchan, M.R.: *Trans* activation of the simian virus 40 enhancer. Mol. Cell. Biol. 6 (1986) 1283–1295.

Schmitz, A. and Galas, D.J.: The interaction of RNA polymerase and *lac* repressor with the *lac* control region. Nucl. Acids Res. 6 (1979) 111–137.

Sellitti, M.A., Pavco, P.A. and Steege, D.A.: *lac* repressor blocks in vivo transcription of *lac* control region DNA. Proc. Natl. Acad. Sci. USA 84 (1987) 3199–3203.

Serfling, E., Jasin, M. and Schaffner, W.: Enhancers and eukaryotic gene transcription. Trends Genet. 1 (1985) 224–230.

Spandidos, D.A. and Wilkie, N.M.: Expression of exogenous DNA in mammalian cells. In Hames, B.D. and Higgins, S.J. (Eds.), Transcription and Translation: A Practical Approach. IRL Press, Oxford, 1984, pp. 1–48.

Spinelli, G. and Ciliberto, G.: Functional activity and chromatin configuration of SV40 enhancer injected in *Xenopus laevis* oocytes. Nucl. Acids Res. 13 (1985) 8065–8081.

von Hippel, P.H., Revzin, A., Gross, C.A. and Wang, A.C.: Non-specific DNA binding of genome regulating proteins as a biological control mechanism, 1. The *lac* operon: equilibrium aspects. Proc. Natl. Acad. Sci. USA 71 (1974) 4808–4812.

Wagner, M.J., Sharp, J.A. and Summers, W.C.: Nucleotide sequence of the thymidine kinase gene of herpes simplex virus type 1. Proc. Natl. Acad. Sci. USA 78 (1981) 1441–1445.

White, M.M., Mayne, K.M., Lester, H.A. and Davidson, N.: Mouse-torpedo hybrid acetylcholine receptors: functional homology does not equal sequence homology. Proc. Natl. Acad. Sci. USA 82 (1985) 4852–4856.

Wickens, M.P. and Gurdon, J.B.: Post-transcriptional processing of simian virus 40 late transcripts in injected frog oocytes. J. M-l. Biol. 163 (1983) 1–26.

Communicated by A.D. Riggs.

CHAPTER 4

INTRODUCTION TO SECTION II

Molecular cloning and characterization of the mouse skeletal α-actin gene.

The actin genes represent a multigene family whose individual members are expressed in a developmentally regulated and tissue-specific manner (Saborio et al., 1979; Vandekerckhove and Weber, 1979; Ordahl et al., 1980; Bruskin et al., 1981; Schwartz and Rothblum, 1981; Minty et al., 1982; Fyrberg et al., 1983; Garcia et al., 1984; Vandekerckhove and Weber, 1984; Nudel et al., 1985). During differentiation of muscle tissue, the β- and γ-cytoskeletal actins are down-regulated, whereas the sarcomeric α-cardiac and α-skeletal actins are up-regulated, synthesized in large amounts, but with distinct tissue-specific and developmental patterns of expression (Devlin and Emerson, 1979; Minty et al., 1982; Gunning et al., 1983; Bains et al., 1984; Mohun et al., 1986; Nudel et al., 1986).

During the terminal development of skeletal muscle cells, proliferating mononucleated myoblasts stop dividing and fuse to form multinucleated myotubes. Coincident with this transition, the relative level of skeletal α-actin mRNA is induced at least 25-fold over its level in perfusion myoblasts both *in vivo* and *in vitro* (Hayward and Schwartz, 1986). It is therefore of particular interest to investigate the molecular mechanisms underlying the selective induction of the skeletal α-actin gene family member in the course of muscle cell development.

What might be the molecular mechanism(s) governing the control of the skeletal α-actin gene regulation in myogenic and nonmyogenic cells? One approach to answering this question is the molecular cloning and characterization of the genes coding for the skeletal α-actin proteins. By studying the structural organization of the skeletal α-actin gene,

we can begin to look for the controlling elements which modulate the expression of this gene during muscle cell development. Hence, we have isolated and characterized the genomic clone of the mouse skeletal α-actin gene from BALB/c mouse sperm DNA.

In this section, we will report on the molecular cloning and characterization of this gene. As described in Chapter 5, the structure and the complete nucleotide sequence of the single genomic copy of the mouse skeletal α-actin gene has been determined. A comparison of the nucleotide sequences of several vertebrate skeletal α-actin genes reveals several highly conserved sequences in the 5′-flanking region and in both 5′- and 3′-untranslated regions. Additionally, two potential inverted repeat sequences, partially within the conserved regions, have been identified within the large first intron of rodent skeletal α-actin genes. It has been reported that a tissue-specific transcriptional enhancer is located in the first intron of a heavy-chain immunoglobulin gene (Gillies et al., 1983). Moreover, a tissue-specific transcriptional enhancer has been identified very recently within the large first intron of the human β-actin gene (Kawamoto et al., 1988). Therefore, in the future it will be interesting to investigate whether or not these partially conserved inverted repeat sequences in the first intron of mammalian skeletal α-actin genes serve as regulatory elements, such as enhancers, in differentiated muscle cells.

In Chapter 6, a new mapping technique for determining transcription start sites on cloned genomic DNA has been developed. We have successfully used this approach to map the transcription start site of the mouse skeletal α-actin gene.

REFERENCES

1. Bains, W., Ponte, P., Blau, H., and Kedes, L. (1984) *Mol. Cell. Biol.* 4, 1449-1453.

2. Bruskin, A.M., Tyner, A.L., Wells, D.E., Showman, R.M.,and Klein, W.H. (1981) *Dev. Biol.* 87, 308-318.

3. Devlin, R.B., and Emerson, C.P. (1979) *Dev. Biol.* 69, 202-216.

4. Fyrberg, E.A., Mahaffey, J.W., Bond, B.J., and Davidson, N. (1983) *Cell* 33, 115-123.

5. Garcia, R., Paz-Aliaga, B., Ernst, S.G., and Crain, Jr., W.R. (1984) *Mol. Cell. Biol.* 4, 840-845.

6. Gillies, S.D., Morrison, S.L., Oi, V.T., and Tonegawa, S. (1983) *Cell* 33, 717-728.

7. Gunning, P., Ponte, P., Blau, H., and Kedes, L. (1983) *Mol. Cell. Biol.* 3, 1985-1995.

8. Hayward, L.J., and Schwartz, R.J. (1986) *J. Cell. Biol.* 101, 1485-1493.

9. Kawanoto, T., Makino, K., Niwa, H., Sugiyama, H., Kimura, S., Amemura, M., Nakata, A., and Kakunaga, T. (1988) *Mol. Cell. Biol.* 8, 267-272.

10. Minty, A.J., Alonso, S., Caravatti, M., and Bunckingham, M.E. (1982) *Cell* 30, 185-192.

11. Mohun, T., Garret, N., and Gurdon, J.B. (1986) *EMBO J.* 5, 3185-3193.

12. Nudel, U., Greenberg, D., Ordahl, C.P., Saxel, O., Newman, S., and Yaffe, D. (1985) *Proc. Natl. Acad. Sci. USA* 82, 3106-3109.

13. Nudel, U., Melloul, D., Greenberg, D., and Yaffe, D. (1986) In

Emerson, C.P., Fischman, D., Nadal-Ginard, B., and Siddiqui, M.A. (eds.) *Molecular biology of muscle development*. Alan R. Liss, Inc., New York, pp. 559-573.

14. Ordahl, C.P., Tilghman, S.M., Ovitt, C., Fornwald, J., Largen, M.T. (1980) *Nucleic Acids Res.* 8, 4989-5005.
15. Saborio, J.S., Segura, M., Flores, M., Garcia, R., and Palmer, E. (1979) *J. Biol. Chem.* 254, 11119-11125.
16. Schwartz, R.J., and Rothblum, K.N. (1981) *Biochemistry* 20, 4122-4129.
17. Vandekerckhove, J., and Weber, K. (1979) *Differentiation* 14, 123-133.
18. Vandekerckhove, J., and Weber, K. (1984) *J. Mol. Biol.* 179, 391-413.

CHAPTER 5

The complete sequence of the mouse skeletal α-actin gene reveals several conserved and inverted repeat sequences outside of the protein-coding region.

Mickey C-T. Hu, Sandra B. Sharp, and Norman Davidson

Molecular and Cellular Biology

6:15-25 (1986)

MOLECULAR AND CELLULAR BIOLOGY, Jan. 1986, p. 15–25
0270-7306/86/010015-11$02.00/0

Vol. 6, No. 1

The Complete Sequence of the Mouse Skeletal α-Actin Gene Reveals Several Conserved and Inverted Repeat Sequences Outside of the Protein-Coding Region

MICKEY CHIEN-TSUNG HU, SANDRA B. SHARP, AND NORMAN DAVIDSON*
Department of Chemistry, California Institute of Technology, Pasadena, California 91125

Received 17 June 1985/Accepted 24 September 1985

The complete nucleotide sequence of a genomic clone encoding the mouse skeletal α-actin gene has been determined. This single-copy gene codes for a protein identical in primary sequence to the rabbit skeletal α-actin. It has a large intron in the 5′-untranslated region 12 nucleotides upstream from the initiator ATG and five small introns in the coding region at codons specifying amino acids 41/42, 150, 204, 267, and 327/328. These intron positions are identical to those for the corresponding genes of chickens and rats. Similar to other skeletal α-actin genes, the nucleotide sequence codes for two amino acids, Met-Cys, preceding the known N-terminal Asp of the mature protein. Comparison of the nucleotide sequences of rat, mouse, chicken, and human skeletal muscle α-actin genes reveals conserved sequences (some not previously noted) outside of the protein-coding region. Furthermore, several inverted repeat sequences, partially within these conserved regions, have been identified. These sequences are not present in the vertebrate cytoskeletal β-actin genes. The strong conservation of the inverted repeat sequences suggests that they may have a role in the tissue-specific expression of skeletal α-actin genes.

The actins represent a multigene family of highly conserved proteins found in all eucaryotes. Differences in amino acid sequence among the various actins have shown that at least six different isoforms are expressed in vertebrates (52, 53). Two striated muscle isoforms, skeletal α and cardiac α (52), and two smooth muscle isoforms (53) are found in the contractile apparatus of muscle fibers, whereas two cytoskeletal isoforms, β and γ, are present in the cytoskeleton of all cells (51). These actin proteins are extremely conserved in amino acid sequence.

Actin gene expression is tissue specific and developmentally regulated (27, 29, 32, 35). By studying the structural organization of the actin gene family, one can begin to look for the controlling elements which modulate the expression of these genes during development. Here we present the complete nucleotide sequence of the single genomic copy of the mouse skeletal α-actin gene. The coding region of this gene is interrupted by five introns which are located in the same positions as introns previously identified in other vertebrate skeletal α-actin genes (13, 54). A comparison of the nucleotide sequences of several vertebrate skeletal α-actin genes reveals several blocks of highly conserved sequences in the 5′-flanking region and in both the 5′- and 3′-untranslated regions. Interestingly, the conserved sequences in the 5′-flanking region and within the first untranslated exon can potentially form several hairpin loops by base pairing between adjacent inverted complementary sequences. These regions do not correspond to the potential hairpin structure in the corresponding portion of the rat cytoskeletal β-actin gene (33). Furthermore, it is possible to form long hairpin loops within the first intron and one stem loop in the 3′-untranslated region upstream from the putative polyadenylation signal ATTAAA. These interesting secondary structures are apparently not present in the vertebrate cytoskeletal β-actin genes. To our knowledge, this is the first description of potential secondary structures in the first intron and among the highly conserved sequences in the 5′-flanking region and both the 5′- and 3′-untranslated regions of vertebrate skeletal α-actin genes. Since the species compared (avian and mammalian) have been separated for more than 250 million years (12), these results indicate a strong selective constraint to conserve these sequences and suggest that they may have an important role in the tissue-specific expression of the skeletal α-actin genes.

* Corresponding author.

MATERIALS AND METHODS

Materials. Restriction endonucleases, T4 DNA ligase, T4 polynucleotide kinase, *Escherichia coli* exonuclease VII, and *E. coli* DNA polymerase I large fragment (Klenow) were purchased from Bethesda Research Laboratories, Boehringer Mannheim Biochemicals, or New England BioLabs. Avian myeloblastosis virus reverse transcriptase was obtained from Life Sciences, Inc. Sp6 RNA polymerase and placental RNasin were obtained from Promega Biotec. Radioactively labeled nucleotides were purchased from Amersham Corp. or New England Nuclear Corp. Unlabeled nucleotides were obtained from P-L Biochemicals, Inc. Synthetic oligonucleotides were synthesized by S. Horvath, Caltech, on an automated DNA synthesizer (21) and purified by electrophoresis through a 20% polyacrylamide–8 M urea preparative gel in Tris borate-EDTA buffer. BALB/c genomic DNA was provided by T. Hunkapiller. The *Drosophila* actin genomic clone DmA2 (Dm5C in reference 14) and a 3′-untranslated region of rat skeletal α-actin cDNA (15) were provided by B. Bond and L. Garfinkel, respectively.

Isolation of genomic actin clones and restriction mapping. A cosmid genomic library of BALB/c mouse sperm DNA, constructed and kindly provided by M. Steinmetz at Caltech, was screened by colony hybridization (49) by using an actin-coding region probe isolated from the *Drosophila* actin genomic clone DmA2 (Dm5C). From 16 positive clones, 1 was tentatively identified to contain the skeletal α-actin gene by hybridization with the conserved (29), isotype-specific (37), 3′-untranslated region of a rat skeletal α-actin cDNA.

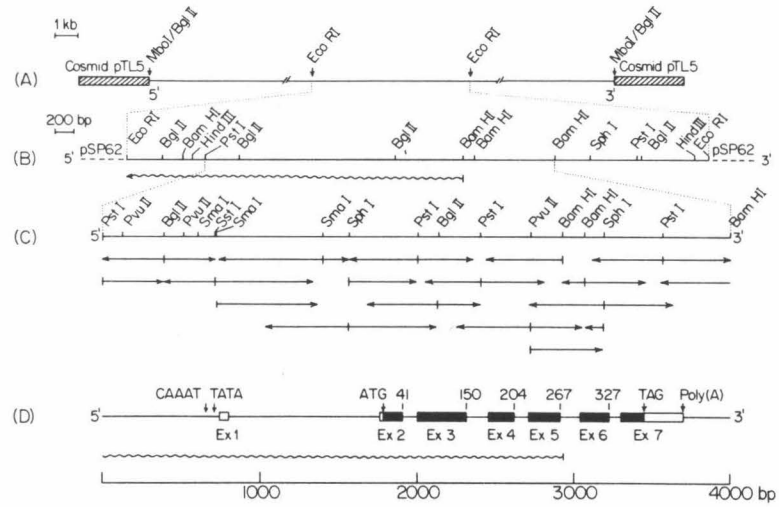


FIG. 1. Structure of the mouse skeletal α-actin gene. (A) Cosmid clone containing the genomic DNA encoding the mouse skeletal α-actin gene. (B) Restriction endonuclease map of the 6.8-kb *Eco*RI DNA fragment. One squiggly arrow represents an SP6 antisense transcript of a *Bam*HI-*Eco*RI fragment that was subcloned from the 6.8-kb *Eco*RI fragment. (C) Detailed restriction map of the mouse skeletal α-actin gene and flanking DNA. Fragments were subcloned into M13 vectors and sequenced by dideoxy chain termination as indicated by arrows. (D) Schematic representation of the structure of the mouse skeletal α-actin gene. Solid boxes represent coding exons, open boxes indicate transcribed untranslated regions, and solid lines coincide with introns and flanking DNA. Numbers above the exons correspond to codon positions (Fig. 2). The squiggly line represents the SP6 antisense transcript as mentioned above.

Southern blot (48) analysis with the *Drosophila* actin probe localized the mouse α-actin-coding region to a single 6.8-kilobase (kb) *Eco*RI fragment in the cosmid clone. This fragment was subcloned into the *Eco*RI site of plasmid pSP62-PL (28), provided by D. Melton, Harvard University.

The restriction endonuclease map of the 6.8-kb *Eco*RI fragment was determined by single and double enzyme digests. Subsequently, the 3.9-kb *Bam*HI-*Eco*RI fragment was subcloned into the *Bam*HI and *Eco*RI sites of the plasmid pSP62-PL (see Fig. 1B) and mapped in finer detail by digestion with more restriction endonucleases.

Localization of the promoter region of the α-actin gene. Several different restriction endonuclease digests of the 6.8-kb *Eco*RI fragment were probed with a 20-base oligonucleotide (5′-GCCCAACACCCAAATATGGC-3′) containing the sequence of the CAAT promoter homology, highly conserved between the skeletal α-actin genes of chickens and rats (34). The oligonucleotide was 5′ end labeled with polynucleotide kinase and [γ-^{32}P]ATP, and hybridization was performed directly in the dried agarose gel as described previously (39). A linearized chicken skeletal α-actin genomic clone (provided by C. Ordahl [13]) and *Eco*RI-linearized SP6 vector were used as positive and negative controls, respectively.

M13 cloning and DNA sequencing. Appropriate restriction fragments from the 6.8- and 3.9-kb inserts were subcloned, in opposite orientations, into the multiple cloning sites of M13 mp18 and M13 mp19 RF vectors, transforming first into *E. coli* JM101 for high efficiency and replating with *E. coli* JM109 (*recA*$^-$) to prevent sequence changes.

Single-stranded M13 templates were sequenced by the dideoxy chain termination procedure (41) with an [α-^{35}S]dATP (500 Ci/mmol) label as described by Biggin et al. (3), with the following modifications. (i) The synthetic pentadecanucleotide (5′-TCCCAGTCACGACGT-3′) and the hexadecanucleotide (5′-GGGTAACGCCAGGGTT-3′) were used as sequencing primers. (ii) The dideoxy sequence reactions were carried out in 50 mM NaCl–7 mM Tris hydrochloride (pH 7.4)–10 mM $MgCl_2$–3 mM dithiothreitol. (iii) The final concentrations of unlabeled nucleotides in each sequence reaction were as follows: A reaction, 25 μM dCTP, 25 μM dGTP, 25 μM dTTP, 20 μM ddATP; C reaction, 8 μM dCTP, 32 μM dGTP, 32 μM dTTP, 50 μM ddCTP; G reaction, 32 μM dCTP, 8 μM dGTP, 32 μM dTTP, 50 μM ddGTP; T reaction, 32 μM dCTP, 32 μM dGTP, 8 μM dTTP, 50 μM ddTTP. (iv) After electrophoresis, the gel was immediately dried, without fixing, for 1 h at 80°C and autoradiographed.

Computer analysis of sequence homology was done as described by Hunkapiller et al. (21).

Primer extension analysis. Polyadenylated [poly(A)$^+$] RNA from a differentiated culture of the mouse myogenic cell line BC3H-1 (42) was isolated by guanidine thiocyanate extraction (9) and two cycles of oligo(dT)-cellulose chromatography (1). A synthetic 42-base oligonucleotide (5′ - AGAGCCGTTGTCACACACAAGAGCGGTGGTCTC GTCTTCGTC-3′) complementary to a portion of the coding sequence, spanning positions 1038 through 1079 in exon 2, was 5′ end labeled with polynucleotide kinase and [γ-^{32}P]ATP (3,000 Ci/mmol) and used as an extension primer. One picomole of the labeled oligonucleotide (10^6 cpm/μg) was denatured by heating at 80°C for 10 min in 40 μl of 98% formamide containing 1 μl of 0.5 M EDTA. Five micrograms of poly(A)$^+$ RNA in 10 μl of 200 mM sodium piperazine-*N*,*N*′-bis(2-ethanesulfonic acid) (pH 6.4)–2 M NaCl–5 mM EDTA was added, and the mixture was incubated at 37°C for 12 h. The nucleic acids were ethanol precipitated from ammonium acetate and reprecipitated from sodium acetate. The dried pellet was suspended in 25 μl of 100 mM Trishydrochloride (pH 8.3)–40 mM KCl–20 mM $MgCl_2$–10

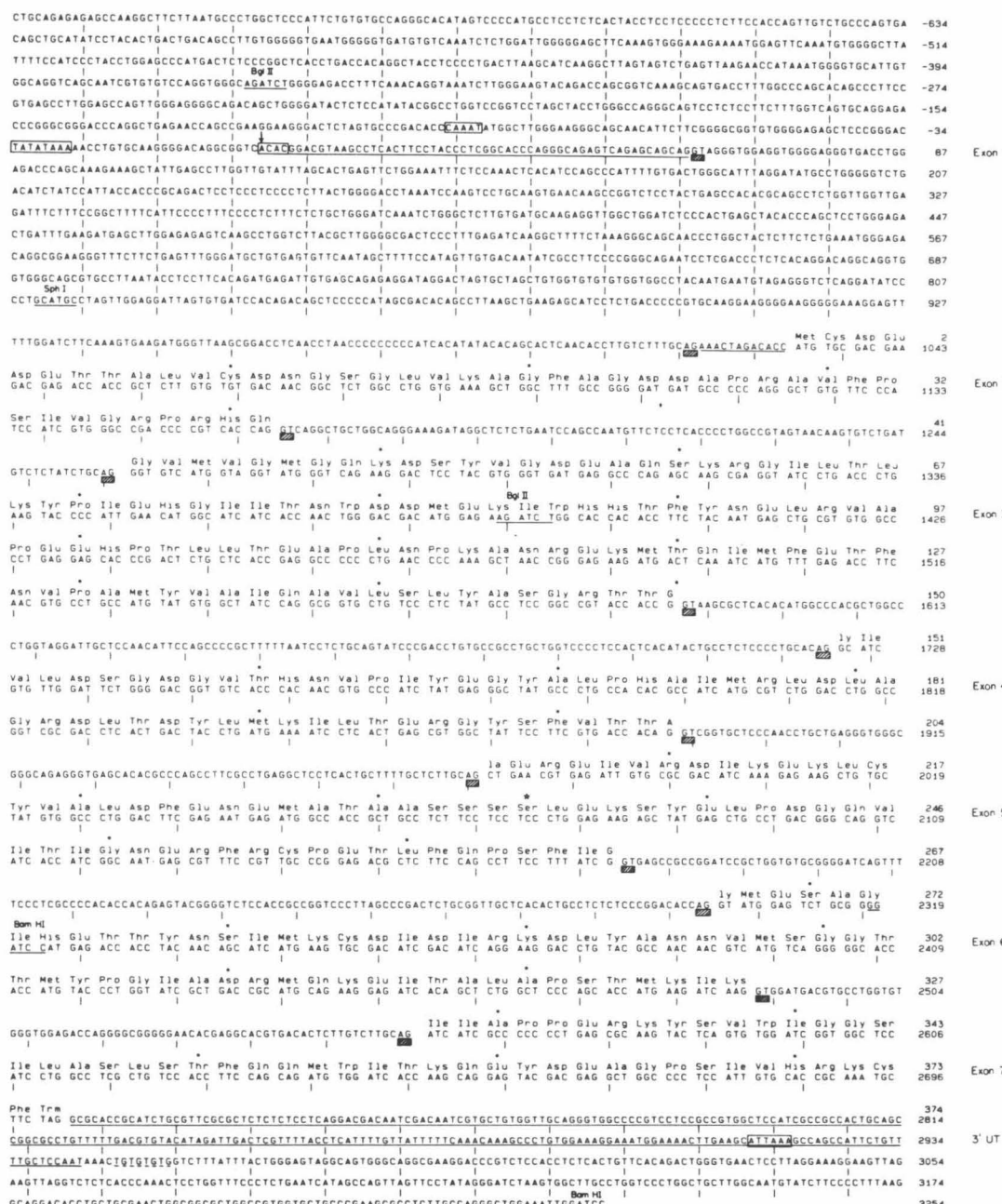


FIG. 2. Nucleotide sequence of the mouse skeletal α-actin gene. Numbers in the left and right margins, respectively, refer to the first and last nucleotides or amino acids in each line. Negative numbers indicate nucleotide positions upstream from the transcription start site. The deduced amino acid sequences encoded by the exons are indicated above the nucleotide sequence in the three-letter amino acid code and numbered as described previously (23, 52). An asterisk indicates an "extra" serine residue between codons 234 and 235, which has been designated position 234a (23, 52). A hatched box represents a splice junction border, and a vertical arrow denotes the cap site which is assigned as nucleotide number 1. The CAAAT, TATA, and putative polyadenylation signal ATTAAA are indicated by the boxes. The 5′- and 3′-untranslated gene regions are underlined, and a G+T-rich stretch downstream from the putative polyadenylation site (16, 26) is underscored with a broken line. The restriction sites used for the exonuclease VII mapping described in the legend to Fig. 3 are marked and underlined. Abbreviations: Trm, termination codon; 3′UT, 3′-untranslated region.

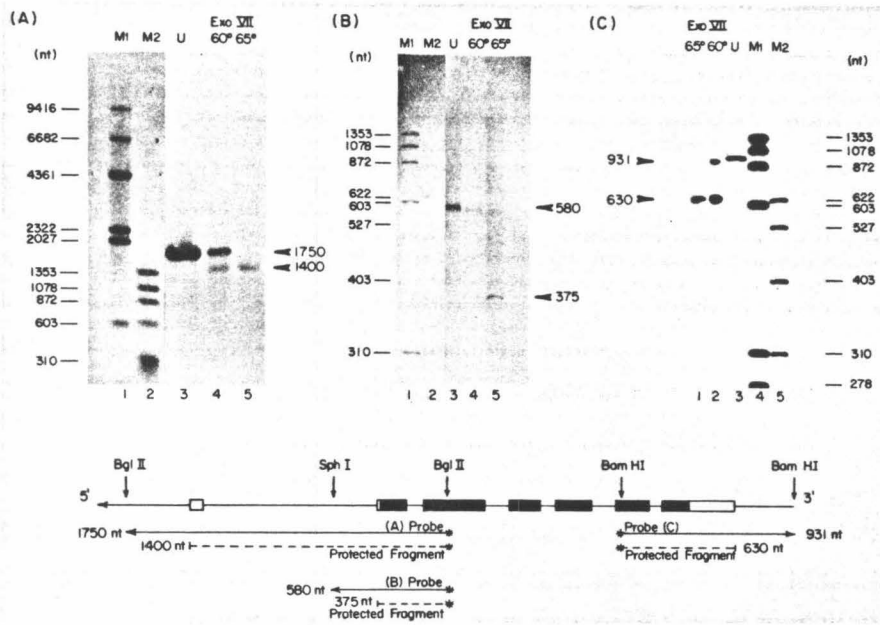


FIG. 3. Exonuclease VII mapping of the mouse skeletal α-actin gene. Total RNA (50 μg) from a differentiated culture of BC3H-1 cells was hybridized at 60 or 65°C with approximately 5 ng of each probe as shown in the schematic diagram (lower panel). The hybridized samples were treated with 3 U of exonuclease VII (1 h, 45°C), electrophoresed on an alkaline agarose gel (A) or a 5% polyacrylamide–8 M urea sequencing gel (B, C), and autoradiographed. (A) Lanes: 1, size markers, *Hin*dIII fragments of γ DNA; 2, size markers, *Hae*III fragments of φX174 DNA; 3, undigested probe of 1,750-nucleotide *Bgl*II-*Bgl*II fragment; 4 and 5, protected products from samples hybridized at 60 and 65°C, respectively. (B) Lanes: 1, size markers, *Hae*III fragments of φX174 DNA; 2, size markers, *Hpa*II fragments of PBR 322; 3, undigested probe of 580-nucleotide *Sph*I-*Bgl*II fragment; 4 and 5, protected products from samples hybridized at 60 and 65°C, respectively. (C) Lanes: 1 and 2, protected products from samples hybridized at 65 and 60°C, respectively; 3, undigested probe of 391-nucleotide *Bam*HI-*Bam*HI fragment; 4, size markers, *Hae*III fragments of φX174 DNA; 5, size markers, *Hpa*II fragments of PBR 322. nt, Nucleotide.

mM dithiothreitol and heated to 65°C for 10 min. This heating step substantially improved the resolution of extension products over background, as observed by Fornwald et al. (13). An equal volume of a solution containing deoxynucleoside triphosphates (1 mM each), 750 U of RNasin per ml, and 40 μg of actinomycin D per ml was added, and the primer extended with avian myeloblastosis virus reverse transcriptase (500 U/ml) at 42°C for 1 h. The reaction was terminated by the addition of EDTA to 10 mM, and the RNA was degraded by treatment with DNase-free RNase (50 μg/ml) at 40°C for 1 h. After ethanol precipitation, the reactions were suspended in formamide-dye buffer, denatured by boiling for 3 min, and electrophoresed on a 6% polyacrylamide–8 M urea sequencing gel in Tris borate-EDTA buffer. DNA sequencing reactions were used as size markers. Extended products were detected by autoradiography of the dried gel.

Exonuclease VII mapping. For exonuclease VII mapping, actin-gene-containing plasmid DNA fragments were isolated and labeled with either polynucleotide kinase and [γ-^{32}P]ATP (3,000 Ci/mmol) for 5′-end mapping, or *E. coli* DNA polymerase I large fragment (Klenow) and [α^{32}P]dCTP (400 Ci/mmol) for 3′-end mapping.

DNA-RNA hybridization was performed as described above in the primer extension analysis by using 50 μg of total RNA from a differentiated culture of BC3H-1 cell line and about 5 ng of labeled (ca. 10^6 cpm/μg) DNA fragment. Hybridizations were carried out at 60 or 65°C for 3 h. Each hybridization mixture was diluted into 10 volumes of 30 mM KCl–10 mM Tris hydrochloride (pH 7.4)–10 mM EDTA, chilled on ice, and incubated at 45°C for 1 h with 10 U of *E. coli* exonuclease VII per ml (2). After ethanol precipitation, the exonuclease VII-resistant material was electrophoresed on a 5% polyacrylamide–8 M urea sequencing gel or alkaline agarose gel (24) and autoradiographed.

In vitro transcription with SP6 RNA polymerase and RNase mapping. Synthesis of the complementary-strand SP6 probe and RNase mapping were carried out as described by Melton et al. (28).

RESULTS

Isolation and mapping of the mouse skeletal α-actin gene. A cosmid clone containing the skeletal α-actin gene was isolated from a BALB/c genomic cosmid library as described in Materials and Methods. The location and orientation of the gene within a single 6.8-kb *Eco*RI fragment were established by restriction endonuclease mapping with 5′ and 3′ fragments of the *Drosophila* actin-coding sequence and the rat α-actin 3′-untranslated sequence as probes.

The 6.8-kb *Eco*RI fragment containing the entire skeletal α-actin-coding and flanking sequences was isolated from the 41-kb insert of the cosmid clone (Fig. 1A) and subcloned into the *Eco*RI site of the plasmid vector pSP62 (Fig. 1B). Subsequently, the putative promoter region of the gene was

localized to about 100 base pairs within the *Sma*I-*Sst*I region (Fig. 1C and D) by hybridization with the 20-base oligonucleotide probe containing the sequence of the CAAT promoter homology which is highly conserved between the skeletal α-actin genes of chickens and rats (34) (data not shown). This result confirmed that the subcloned *Eco*RI fragment contained the 5′-flanking sequence of the gene. Also, it suggested that the position of the transcription initiation site should be close to the *Sst*I site. The detailed restriction endonuclease map of the actin-coding and flanking regions in Fig. 1C was used to choose the DNA fragments to be subcloned into M13 and mp18 and mp19 vectors for sequencing. From these subclones we determined the linear sequence of 4,007 nucleotides, on both strands independently (Fig. 2).

Amino-acid-coding region of the mouse skeletal α-actin gene. The complete nucleotide sequence of the mouse skeletal α-actin gene with the 5′- and 3′-flanking regions is shown in Fig. 2. Exons within the protein-coding region and the introns separating them were initially assigned mainly by homology with the rat skeletal α-actin genomic sequence (54) and in part by comparison with the partial cDNA sequence for the carboxy-terminal portion of the mouse protein (29). This assignment was supported by the fact that the sequences at the presumed exon-intron junctions are in accordance with the consensus sequence for splice sites (7). In addition, the expected lengths of exons 2, 3, 4, and 5 were confirmed experimentally within an accuracy of ±3 nucleotides by hybridizing an SP6 anti-sense transcript of a *Bam*HI-*Eco*RI fragment (Fig. 1B) to poly(A)$^+$ RNA from differentiated BC3H-1 cells and RNase mapping (data to be presented in the Ph.D thesis by M. C.-T. Hu at the California Institute of Technology). The translated amino acid sequence for this interpretation of the structure of the mouse skeletal α-actin gene is identical to that of rats, rabbits, and chickens. (The amino acid sequence of mouse skeletal α-actin has not been directly determined.)

The coding sequence begins with codons for two amino acids, Met and Cys, which are absent from the mature protein. They are followed by the codon for Asp (GAC), the known N-terminal residue of striated muscle actin. The same two codons preceding the codon specifying the N-terminal amino acids are found in the human (20), rat (54), and chicken (13) skeletal α-actin genes and in the human (19) and chicken (8) cardiac α-actin genes. Interestingly, these two codons are also found in all six *Drosophila* actin genes (14) and sea urchin actin genes (10, 43), but the Cys codon is absent in vertebrate cytoskeletal β-actin genes (22, 33, 38). It has been suggested that the Met-Cys dipeptide is removed by posttranslational processing (54).

Although the derived sequence of the primary skeletal α-actin is 377 residues, we have numbered the amino acids in Fig. 2 in conformity with the numbering system suggested by Lu and Elzinga (23) and Vandekerckhove and Weber (52), which yields 374 positions. Of the three additional positions, two are the Met-Cys dipeptide at the N-terminus, while the third is an "extra" serine residue between positions 234 and 235 which has been designated 234a (23, 52).

The coding region of the mouse skeletal α-actin gene is split by five introns (Fig. 1D) at codons specifying amino acids 41/42 (IVS 2), 150 (IVS 3), 204 (IVS 4), 267 (IVS 5), and 327/328 (IVS 6). These intron positions are identical to those for the corresponding genes of chickens and rats (13, 54). The length and positions of the exons and introns may be deduced from Fig. 2. Previously, Zakut et al. (54) have reported that a potential splice site (CAG/GTA) is present 32

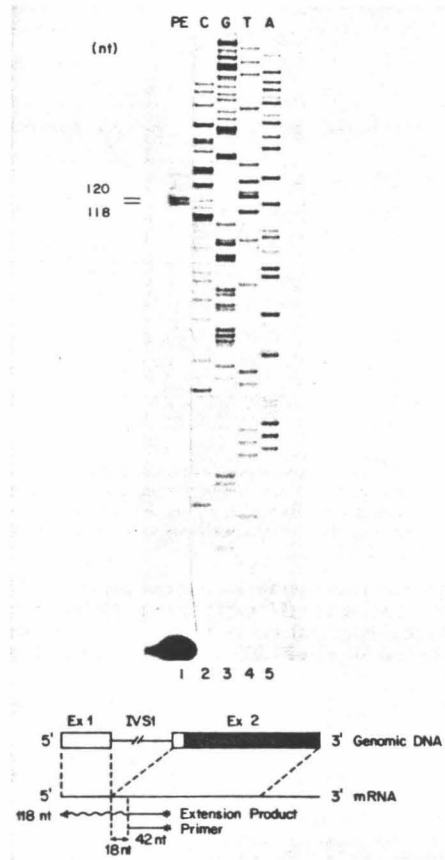


FIG. 4. Identification of the 5′ end of the mouse skeletal α-actin mRNA by primer extension. The 5′-end-labeled, 42-base oligonucleotide complementary to the coding sequence, from positions 1038 to 1079 (Fig. 2), was hybridized with 5 μg of poly(A)$^+$ RNA from a differentiated culture of BC3H-1 cells and extended by using reverse transcriptase as described in the text. The extension products were electrophoresed on a 6% polyacrylamide–8 M urea sequencing gel and autoradiographed. The diagram (lower panel) shows the product expected from full-length elongation of mRNA. Lanes: 1, primer extension products; 2 through 5, sequencing ladders used as size markers. nt, Nucleotide.

base pairs downstream from the CG/GT splice site at codon 150 in the rat skeletal α-actin gene. Our results do not reveal this potential splice site in the mouse gene. This is not surprising, because use of the extra splice site in the rat gene would produce an actin with an insert of 11 amino acids, and no such product has yet been detected.

Sequence of the 5′-untranslated region of the mouse skeletal α-actin gene. Although the actin amino acid sequence data and cloned cDNA partial sequence (29) could be used to identify the translated regions of the gene, independent means were required for delineating the 5′-untranslated region. The 5′ borders of the first, untranslated, exon and the second exon were approximately determined by exonuclease VII mapping (Fig. 3A and B). In addition, the precise assignment of the transcription initiation site, ACAC, was

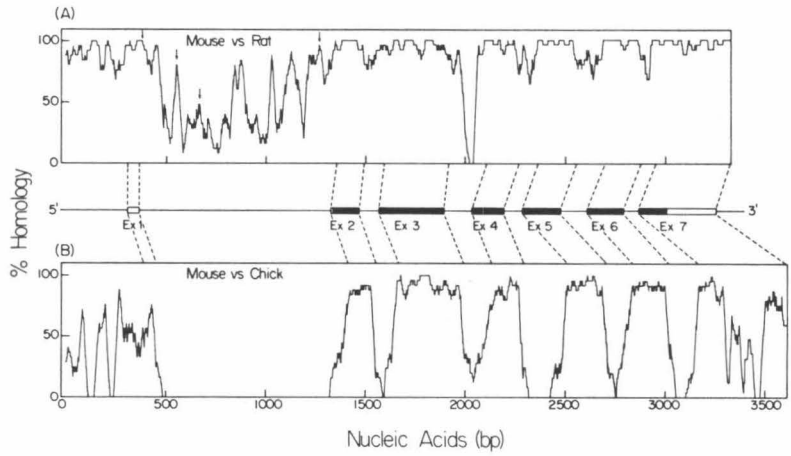


FIG. 5. Percent homology profiles for the mouse, rat, and chicken skeletal α-actin genes. In this analysis, the sequence search string was 25 nucleotides, and gaps inserted in the sequences for alignment purposes were scored as regions of 0% homology. (A) Mouse versus rat skeletal α-actin genes; (B) mouse versus chicken skeletal α-actin genes. The structure of the mouse skeletal α-actin gene is shown between the two homology plots. Vertical arrows indicate two inverted duplications as described in the text. bp, Base pairs.

confirmed by a newly developed mapping technique by using T4 DNA polymerase (M. C. T. Hu and N. Davidson, submitted for publication) and assigned to be nucleotides −1/1. This site was identified 1,031 nucleotides upstream from the initiator ATG codon (Fig. 2). We assigned the 5′ border of the first intron to nucleotides 58/59 by matching the length of the primer extended product (Fig. 4) with the positions determined above for the 5′ borders of the first and second

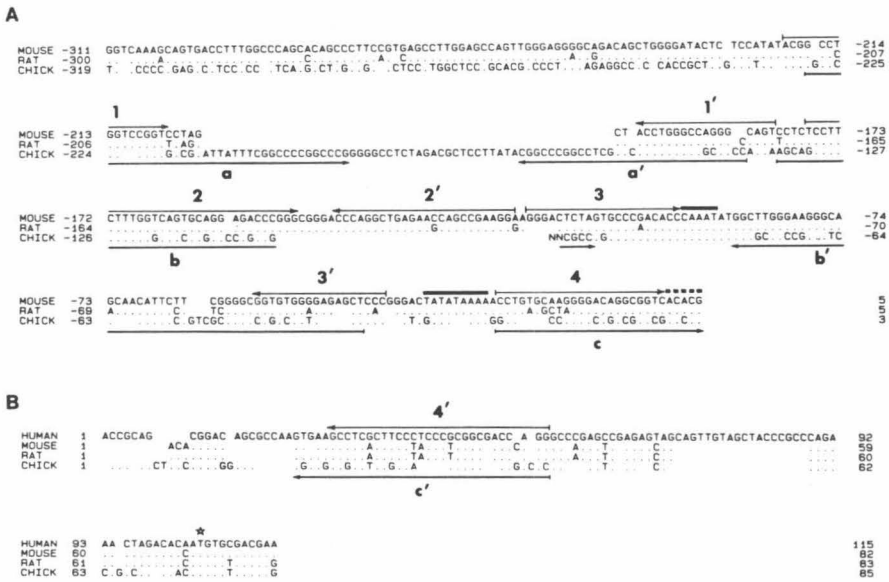


FIG. 6. Comparison of the nucleotide sequence of the 5′-flanking and the 5′-untranslated regions of vertebrate skeletal α-actin genes. (A) Alignment of the 5′-flanking region sequences of mouse, rat, and chicken skeletal α-actin genes. (B) Alignment of the 5′-untranslated region sequences of human, mouse, rat, and chicken skeletal α-actin genes. Dots indicate identity with the first sequence listed. Blanks indicate that gaps have been introduced during the alignment for maximal homology. The CAAAT box and TATA box are highlighted with solid bars. A broken bar indicates the transcription initiation site, and an asterisk indicates the initiation codon ATG. Horizontal arrows above the sequences represent the adjacent inverted complementary sequences of rodents which are indicated numerically, whereas horizontal arrows underneath the sequences represent the adjacent repeats of chicken sequences, which are indicated alphabetically.

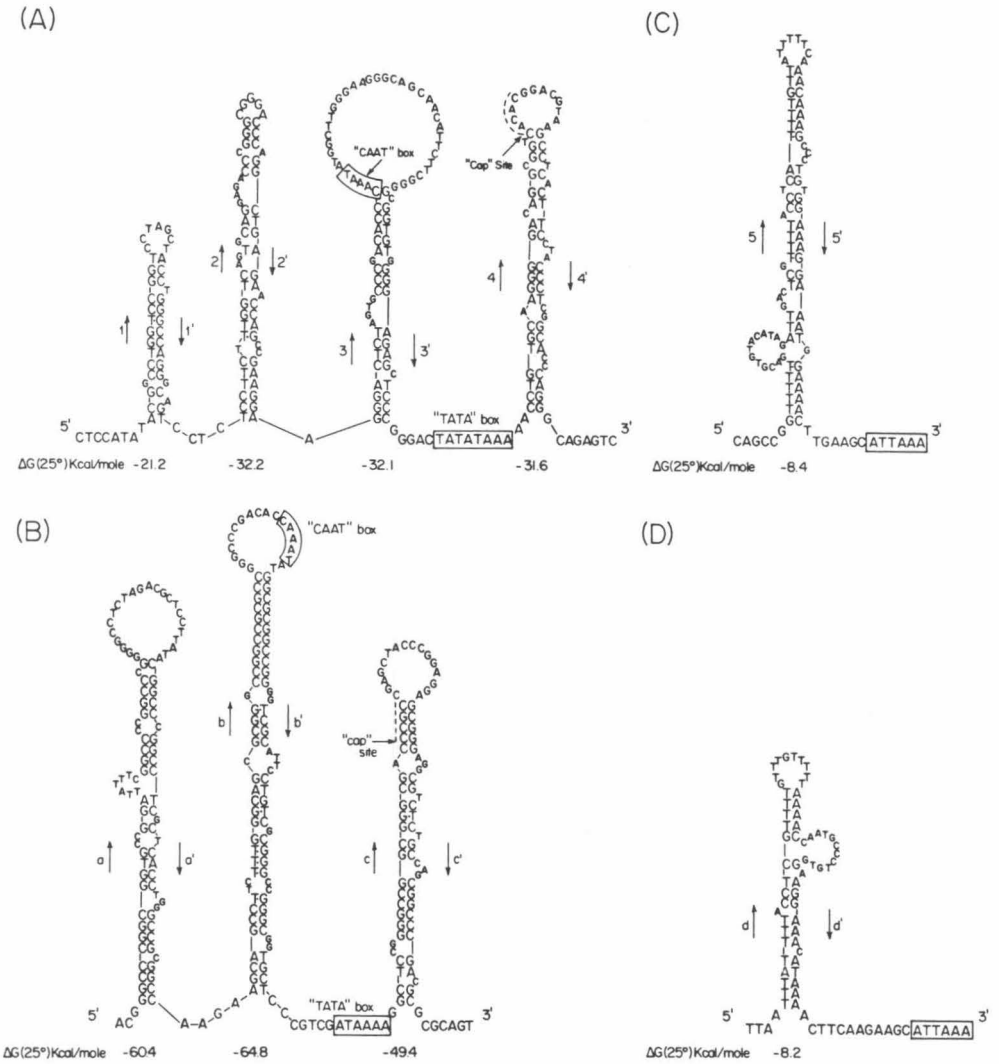


FIG. 7. Predicted inverted repeat structures in the 5′-flanking region and the 5′- and 3′-untranslated regions of the mouse, rat, and chicken skeletal α-actin genes. (A) Four potential configurations are shown in the 5′-flanking and the 5′-untranslated region of the mouse skeletal α-actin gene. Similar configurations can be found in the rat gene. (B) Three potential configurations are shown in the 5′-flanking and the 5′-untranslated region of the chicken skeletal α-actin gene. (C) A potential stem-loop is demonstrated in the 3′-untranslated region of the mouse skeletal α-actin gene. A similar structure can be found in the rat gene. (D) A potential stem-loop is shown in the 3′-untranslated region of the chicken skeletal α-actin gene. The indicated free energy values for the base-paired regions were calculated by the method of Tinoco et al. (50). Note that the free energies for the base-paired regions in the 5′-flanking region were estimated by the same method, assuming that the stacking energies of DNA base pairs are similar to those of RNA base pairs. CAAAT, TATA, and putative polyadenylation signal ATTAAA are indicated by the boxes.

exons. The sequences at the determined borders of the first intron are in agreement with the consensus splice site sequences (7). Thus, the first intron is 961 nucleotides long and interrupts the 5′-untranslated region 12 nucleotides upstream from the initiator ATG codon.

A canonical promoter sequence TATATAAA (5) was identified at nucleotides −33 to −26 (Fig. 2), and a CAAAT sequence (12) was located 91 nucleotides upstream from the transcription initiation site. The positions of these regulatory sequences in the promoter region correspond well with those of similar sequences found upstream from the 5′ cap site of other eucaryotic genes (7).

Sequence of the 3′-untranslated region of the mouse skeletal α-actin gene. The location of the polyadenylation site was

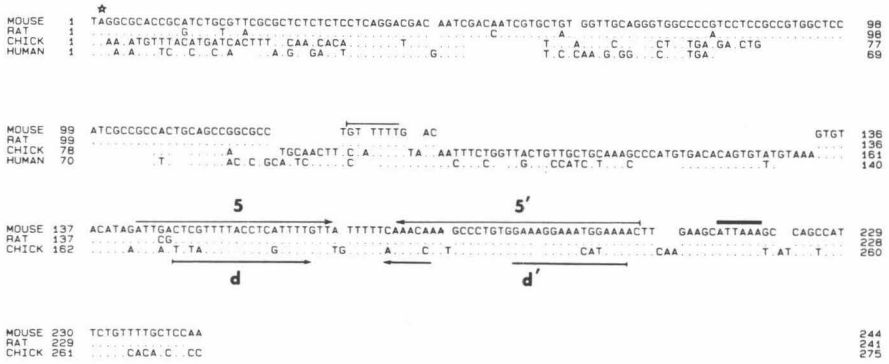


FIG. 8. Comparison of the nucleotide sequence of the 3'-untranslated regions of vertebrate skeletal α-actin genes. Alignment of the 3'-untranslated region sequences of mouse, rat, chicken, and human (partial) skeletal α-actin genes. Symbols are as described in the legend to Fig. 6. The putative polyadenylation signal ATTAAA is highlighted with a solid bar, and the termination codon TAG is indicated with an asterisk.

identified approximately at nucleotide 2944 (±5 nucleotides) of the sequenced gene (Fig. 2) by exonuclease VII mapping (Fig. 3C). The sequence ATTAAA was located 25 nucleotides upstream from the polyadenylation site. In addition, the sequence TGTGTGTGG was found 4 nucleotides downstream from the polyadenylation site, in agreement with the suggestion that a G+T-rich stretch downstream from the polyadenylation site is required for the correct 3'-end formation of mRNA (16, 26). Another potential polyadenylation signal AATAAA was also found at nucleotides 2942 through 2947, about 22 nucleotides downstream from the end of the putative polyadenylation signal ATTAAA identified above. We have no evidence that this second potential signal functions in the mouse.

Thus, the 3'-untranslated region of the mouse skeletal α-actin mRNA is about 245 nucleotides long [excluding the poly(A) tail]. It is about the same size as the 3'-untranslated regions of the rat (241 nucleotides [54]) and human (253 nucleotides [37]) skeletal α-actin mRNA but much shorter than that of rat cytoskeletal β-actin mRNA (670 nucleotides [33]). The length of skeletal α-actin mRNA in mammals is about 1,650 nucleotides (31, 37, 44), including the poly(A) tail. Comparison with the total length of transcribed sequences in the mouse and rat genes (about 1,450 nucleotides) would suggest that the poly(A) tail is about 200 nucleotides long.

Copy number of the mouse skeletal α-actin gene. Southern blot analysis of BALB/c genomic DNA digested with four different restriction endonucleases (*Pst*I, *Bgl*II, *Hin*dIII, and *Sst*I) demonstrated that genomic and cosmid fragments hybridizing with the skeletal α 3'-untranslated region-specific probe comigrate (data not shown). This result suggests that the gene is present in single copy in the mouse genome, in agreement with previous reports (29, 30).

DISCUSSION

Strong homology and interesting inverted repeat structures in the 5'-flanking region and both the 5'- and 3'-untranslated regions of vertebrate skeletal α-actin genes. We have aligned the nucleotide sequence of the mouse skeletal α-actin gene with those of the rat and chicken by using percent homology profiles (Fig. 5). All alignments depend on introduction of gaps for maximal homology, and areas of high and low homology between two sequences are displayed as peaks and troughs, respectively. The coding sequences show a very high degree of homology (≥90%), as expected since the proteins themselves are identical. In comparison of rat with mouse, the intron sequences between coding exons are about 75% homologous, except for intron 3, for which the lengths differ by 49 nucleotides. The corresponding introns of the chicken are much more divergent in length and sequence. In comparing rat with mouse, the long intron following the 5' untranslated exon 1 shows sharp peaks of conserved and nonconserved regions.

Figures 5 and 6A show very high conservation (~85%) in the 5'-flanking region between the cap site and 300 nucleotides upstream of rat and mouse skeletal α-actin genes. The homology between chicken and mouse in the same region is also rather high (~60%). Nudel et al. (32) have found a similar degree of homology between the rat and chicken skeletal α-actin genes. We have also found considerable homology in the 5'-untranslated region. By introducing gaps for best alignment (Fig. 6B) there is (i) a high degree of homology between rat and mouse in the 5' untranslated region, (ii) a rather high degree of homology between human and rodents, except for three long inserts in the human gene, and (iii) a moderate degree of homology between chickens and rodents. Conserved sequences between chickens and rats around the CAAT box and about 46 to 59 nucleotides downstream from the cap site have been previously recognized by Ordahl and Cooper (34). A similar comparison of the 5'-untranslated regions of nonmuscle β-actins (human cDNA and rat genomic sequence) also shows a high degree of sequence conservation (38). We find, however, no cross-homology between α- and β-actin 5'-untranslated regions. A number of studies have suggested that the sequence and structure of the mRNA in the 5'-untranslated region have an important role in regulation of translation (11, 36, 40, 46). The fact that there are conserved sequences in the 5'-untranslated region in all of the vertebrate skeletal α-actin genes, but a different set of conserved sequences in the cytoskeletal β-actin genes, suggests that there may be developmentally specific translational regulatory mechanisms in muscle versus nonmuscle cells.

It is striking that a number of inverted repeat structures exist in the 5'-flanking and the 5'-untranslated regions of the rodent and chicken genes. These are indicated as inverted repeats in Fig. 6 and as remarkably stable hairpin structures

for a single strand in Fig. 7A and B. Some of these inverted repeat structures have been conserved between chickens and rodents. The species compared (avian and mammalian) have been separated for more than 250 million years (12), indicating a strong selective constraint to conserve these sequences and suggesting that the sequences may be biologically significant. If the primary transcripts actually initiate at the cap site, the structures shown in Fig. 7A and B would not occur in the RNA. These sequences could function as duplicated transcription factor binding sites, with the bound factors (presumably proteins) having opposite orientation at the two members of an inverted repeat as postulated by McKnight et al. (25) and by Giniger et al. (18). Alternatively, some single-strand DNA regions may be opened up during formation of a transcription bubble, and these hairpins could then form in the DNA as indicated in Fig. 7. Potential hairpin structures have also been found in the 5′-flanking region and the first untranslated exon in the rat cytoskeletal β-actin gene (33). They differ, however, from those found in the skeletal α-actin genes. For example, whereas the TATA box is presented within a loop of the rat β gene, it is found between stem-loop structures in the α genes. The existence of such differences between the cytoskeletal and muscle-specific actin genes raises the possibility that the conserved inverted repeats in the α genes are important for tissue-specific expression. To our knowledge, this is the first description of inverted repeats and possible secondary structure formation among the highly conserved sequences in the 5′-flanking and the 5′-untranslated regions of vertebrate skeletal α-actin genes.

A long sequence of about 110 nucleotides, including the putative polyadenylation signal ATTAAA, is highly conserved in the 3′-untranslated region of vertebrate skeletal α-actin genes (Fig. 8). It is noteworthy that two blocks of these highly conserved sequences in the 3′-untranslated region can form a stem-loop structure with estimated stabilities of −8.4 kcal/mol for mice and rats (Fig. 7C) and −8.2 kcal/mol for chickens (Fig. 7D). These structures are immediately upstream from the ATTAAA polyadenylation signal. It has been suggested that the inverted repeat at the 3′ end of sea urchin histone mRNA is important for the generation of the histone mRNA 3′ termini (5, 6). This putative structure does not act as a DNA cruciform, but exerts its function at the level of the RNA transcripts (4). It is conceivable that the potential hairpin structure upstream from the putative polyadenylation signal in skeletal α-actin plays a role in the correct 3′-end formation of skeletal α-actin mRNA.

Interesting features in the first intron of vertebrate skeletal α-actin genes. Rat and mouse skeletal α-actin genes have a long first intron compared with chickens (i.e., 976, 961, and 111 nucleotides for rats, mice, and chickens, respectively). There are several highly conserved sequences in this intron between rats and mice, but the introns are quite divergent in other regions (Fig. 5A). The chicken first intron is quite G+C rich (82%) compared with rat introns (53.5%) and mouse introns (52%). We have been unable to find any sequence homology between the chicken and rodent introns. There are two inverted duplications within the rodent first intron. Both occur within the conserved sequences (arrows in Fig. 5A). Furthermore, an inverted repeat can also be found in the chicken first intron (shown as a hairpin in Fig. 9C). The remarkably stable hairpin structures in the vertebrate first intron (Fig. 9) may form in the primary transcript. If the splicing apparatus tracks along the intron in search of splice sites, it may be able to pass along the base of such hairpins. The hairpins would effectively shorten the intron, thereby

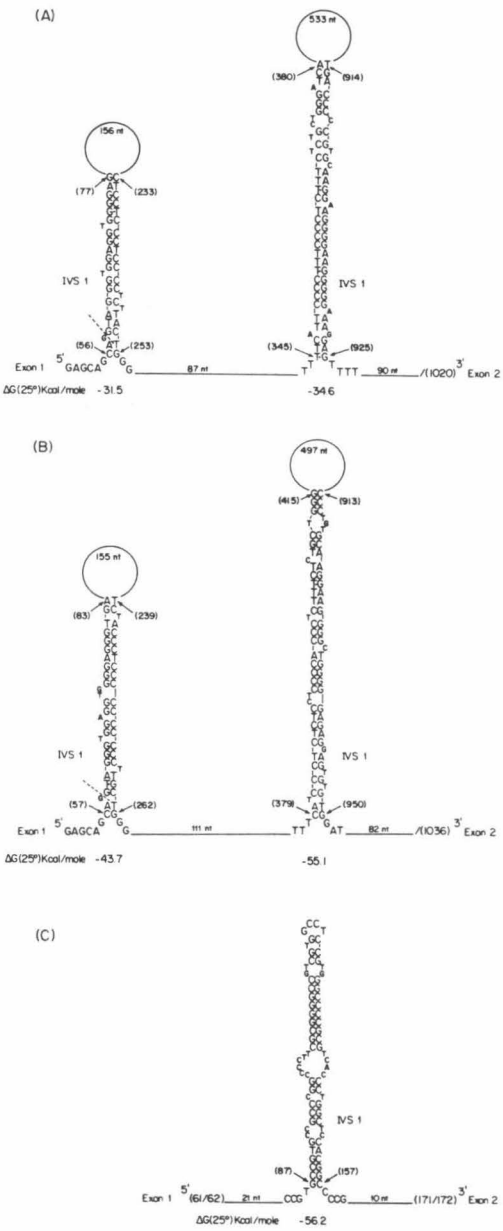


FIG. 9. Predicted inverted repeat structures in the first intron of rodents and chicken skeletal α-actin genes. (A) Two potential configurations in the first intron of mouse skeletal -actin gene. (B) Similar configurations in the first intron of rat skeletal α-actin gene. (C) One potential hairpin loop in the first intron of chicken skeletal α-actin gene. The indicated free energy values for the base-paired regions were calculated by the method of Tinococ et al. (50). Numbers within parentheses indicate the numbers of nucleotides downstream from the transcription initiation site.

expediting its excision by splicing. A similar mechanism has been proposed as one way to explain intermolecular splicing between two RNAs base paired through their intron sequences (47). Alternatively, we speculate that these inverted repeats may play a role of transcriptional enhancement in the regulation of tissue-specific gene expression because it has been proposed that a tissue-specific transcription enhancer element is located in the intron of a heavy-chain immunoglobulin gene (17). In fact, the putative "core" sequence in the heavy-chain gene intron is present as an inverted repeat (17).

In conclusion, we have found several conserved and inverted repeat sequences outside of the protein-coding region of the skeletal α-actin genes. It would be interesting to investigate whether these conserved inverted repeat sequences serve as regulatory elements in differentiated muscle cells.

ACKNOWLEDGMENTS

We are grateful to Michael Steinmetz for generously providing a BALB/c genomic cosmid library, to Charles P. Ordahl for the chicken skeletal α-actin genomic clone, to Leonard I. Garfinkel for a rat α-actin cDNA, to Beverley J. Bond for the *Drosophila* actin DmA2 genomic clone, and to Tim Hunkapiller for BALB/c genomic DNA and help with the computer analyses. We also thank Joan Steitz for valuable discussions and David Solnick, Bruce J. Nicholson, Terry P. Snutch, Robert Little, and Nevis L. Fregien for their helpful comments on this manuscript.

This work was supported by a Public Health Service research grant from the National Institutes of Health.

LITERATURE CITED

1. **Aviv, H., and P. Leder.** 1972. Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid-cellulose. Proc. Natl. Acad. Sci. USA **69**:1408–1412.
2. **Berk, A. J., and P. A. Sharp.** 1978. Spliced early mRNAs of simian virus 40. Proc. Natl. Acad. Sci. USA **75**:1274–1278.
3. **Biggin, M. D., T. J. Gibson, and G. F. Hong.** 1983. Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. Proc. Natl. Acad. Sci. USA **80**:3963–3965.
4. **Birchmeier, C., W. Folk, and M. L. Birnstiel.** 1983. The terminal RNA stem-loop structure and 80 bp of spacer DNA are required for the formation of 3′ termini of sea urchin H2A mRNA. Cell **35**:433–440.
5. **Birchmeier, C., R. Grosschedl, and M. L. Birnstiel.** 1982. Generation of authentic 3′ termini of an H2A mRNA *in vivo* is dependent on a short inverted DNA repeat and on spacer sequences. Cell **28**:739–745.
6. **Birchmeier, C., D. Schumperli, G. Sconzo, and M. L. Birnstiel.** 1984. 3′ editing of mRNAs: sequence requirements and involvement of a 60-nucleotide RNA in maturation of histone mRNA precursors. Proc. Natl. Acad. Sci. USA **81**:1057–1061.
7. **Breathnach, R., and P. Chambon.** 1981. Organization and expression of eucaryotic split genes coding for proteins. Annu. Rev. Biochem. **50**:349–383.
8. **Chang, K. S., K. N. Rothblum, and R. J. Schwartz.** 1985. The complete sequence of the chicken α-cardiac actin gene: a highly conserved vertebrate gene. Nucleic Acids Res. **13**:1223–1237.
9. **Chirgwin, J. M., A. E. Przybyla, R. J. MacDonald, and W. J. Rutter.** 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemstry **18**: 5294–5299.
10. **Cooper, A. D., and W. R. Crain.** 1982. Complete nucleotide sequence of a sea urchin actin gene. Nucleic Acids Res. **10**:4081–4092.
11. **Darveau, A., J. Pelletier, and N. Sonenberg.** 1985. Differential efficiencies of *in vitro* translation of mouse c-myc transcripts differing in the 5′ untranslated region. Proc. Natl. Acad. Sci. USA **82**:2315–2319.
12. **Efstratiadis, A., J. W. Posakony, T. Maniatis, R. M. Lawn, C. O'Connel, R. A. Spritz, J. K. DeRiel, B. G. Forget, S. M. Weissman, J. L. Slightom, A. E. Blechl, O. Smithies, F. E. Baralle, C. C. Shoulders, and N. J. Proudfoot.** 1980. The structure and evolution of the human β-globin gene family. Cell **21**:653–668.
13. **Fornwald, J. A., G. Kuncio, I. Peng, and C. P. Ordahl.** 1982. The complete nucleotide sequence of the chick α-actin gene and its evolutionary relationship to the actin gene family. Nucleic Acids Res. **10**:3861–3876.
14. **Fyrberg, E. A., B. J. Bond, N. D. Hershey, K. S. Mixter, and N. Davidson.** 1981. The actin genes of *Drosophila*: protein coding regions are highly conserved but intron positions are not. Cell **24**:107–116.
15. **Garfinkel, L. I., M. Periasamy, and B. Nadal-Ginarcho.** 1982. Cloning and characterization of cDNA sequences corresponding to myosin light chains 1, 2, and 3, troponin-C, troponin-T, α-tropomyosin, and α-actin. J. Biol. Chem. **257**:11078–11086.
16. **Gil, A., and N. J. Proudfoot.** 1984. A sequence downstream of AAUAAA is required for rabbit β-globin mRNA 3′-end formation. Nature (London) **312**:473–474.
17. **Gillies, S. D., S. L. Morrison, V. T. Oi, and S. Tonegawa.** 1983. A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. Cell **33**:717–728.
18. **Giniger, E., S. M. Varnum, and M. Ptashne.** 1985. Specific DNA binding of GAL4, a positive regulatory protein of yeast. Cell **40**:767–774.
19. **Hamada, H., M. G. Petrino, and T. Kakunaga.** 1982. Molecular structure and evolutionary origin of human cardiac muscle actin gene. Proc. Natl. Acad. Sci. USA **79**:5901–5905.
20. **Hanauer, A., M. Levin, R. Heilig, D. Daegelen, A. Kahn, and J. L. Mandel.** 1983. Isolation and characterization of cDNA clones for human skeletal muscle α-actin. Nucleic Acids Res. **11**:3503–3516.
21. **Hunkapiller, M., S. Kent, M. Caruthers, W. Dreyer, J. Firca, C. Giffin, S. Horvath, T. Hunkapiller, P. Tempst, and L. Hood.** 1984. A microchemical facility for the analysis and synthesis of genes and proteins. Nature (London) **310**:105–111.
22. **Kost, T. A., N. Theodorakis, and S. H. Hughes.** 1983. The nucleotide sequence of the chick cytoplasmic β-actin gene. Nucleic Acids Res. **11**:8287–8301.
23. **Lu, R. C., and M. Elzinga.** 1977. Partial amino acid sequence of brain acin and its homology with muscle actin. Biochemistry **16**:5801–5806.
24. **McDonell, M. W., M. N. Simon, and F. W. Studier.** 1977. Analysis of restriction fragments of T7 DNA and determination of molecular weights by electrophoresis in neutral and alkaline gels. J. Mol. Biol. **110**:119–146.
25. **McKnight, S. L., R. C. Kingsbury, A. Spence, and M. Smith.** 1984. The distal transcription signals of the herpesvirus tk gene share a common hexanucleotide control sequence. Cell **37**: 253–262.
26. **McLauchlan, J., D. Gaffney, J. L. Whitton, and J. B. Clements.** 1985. The consensus sequence YGTGTTYY located downstream from the AATAAA signal is required for efficient formation of mRNA 3′ termini. Nucleic Acids Res. **13**:1347–1368.
27. **Melloul, D., B. Aloni, J. Calvo, D. Yaffe, and U. Nudel.** 1984. Developmentally regulated expression of chimeric genes containing muscle actin DNA sequences in transfected myogenic cells. EMBO J. **3**:983–990.
28. **Melton, D. A., P. A. Krieg, M. R. Rebagliati, T. Manistis, K. Zinn, and M. R. Green.** 1984. Efficient *in vitro* synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. Nucleic Acids Res. **12**:7035–7056.
29. **Minty, A. J., S. Alonso, M. Caravatti, and M. E. Buckingham.** 1982. A fetal skeletal muscle actin mRNA in the mouse and its identity with cardiac actin mRNA. Cell **30**:185–192.
30. **Minty, A. J., S. Alonso, J. L. Guenet, and M. E. Buckingham.** 1983. Number and organization of actin-related sequences in the mouse genome. J. Mol. Biol. **167**:77–101.
31. **Minty, A. J., M. Caravatti, B. Robert, A. Cohen, P. Daubas, A. Weydert, F. Gros, and M. E. Buckingham.** 1981. Mouse actin

messenger RNAs. J. Biol. Chem. **256**:1008–1014.
32. **Nudel, U., D. Greenberg, C. P. Ordahl, O. Saxel, S. Newman, and D. Yaffe.** 1985. Developmentally regulated expression of a chick muscle-specific gene in stably transfected rat myogenic cells. Proc. Natl. Acad. Sci. USA **82**:3106–3109.
33. **Nudel, U., R. Zakut, M. Shani, S. Neuman, Z. Levy, and D. Yaffe.** 1983. The nucleotide sequence of the rat cytoplasmic α-actin gene. Nucleic Acids Res. **11**:1759–1771.
34. **Ordahl, C. P., and T. A. Cooper.** 1983. Strong homology in promoter and 3′-untranslated regions of chick and rat α-actin genes. Nature (London) **303**:348–349.
35. **Ordahl, C. P., S. M. Tilghman, C. Ovitt, J. Fornwald, and M. T. Largen.** 1980. Structure and developmental expression of the chick α-actin gene. Nucleic Acids Res. **8**:4989–5005.
36. **Pelletier, J., and N. Sonenberg.** 1985. Insertion mutagenesis to increase secondary structure within the 5′ noncoding region of a eukaryotic mRNA reduces translational efficiency. Cell **40**: 515–526.
37. **Ponte, P., P. Gunning, H. Blau, and L. Kedes.** 1983. Human actin genes are single copy for α-skeletal and α-cardiac actin but multicopy for β- and γ-cytoskeletal genes: 3′ untranslated regions are isotype specific but are conserved in evolution. Mol. Cell. Biol. **3**:1783–1791.
38. **Ponte, P., S. Y. Ng, J. Engel, P. Gunning, and L. Kedes.** 1984. Evolutionary conservation in the untranslated regions of actin mRNAs: DNA sequence of a human beta-actin cDNA. Nucleic Acids Res. **12**:1687–1696.
39. **Purrello, M., and I. Balazs.** 1983. Direct hybridization of labeled DNA to DNA in agarose gels. Anal. Biochem. **128**:393–397.
40. **Saito, H., A. C. Hayday, K. Wiman, W. S. Hayward, and S. Tonegawa.** 1983. Activation of the c-myc gene by transformation: a model for translational control. Proc. Natl. Acad. Sci. USA **80**:7476–7480.
41. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
42. **Schubert, D., A. J. Harris, C. E. Devine, and S. Heinemann.** 1974. Characterization of a unique muscle cell line. J. Cell Biol. **61**:398–413.
43. **Schuler, M. A., P. McOsker, and E. B. Keller.** 1983. DNA sequence of two linked actin genes of sea urchin. Mol. Cell. Biol. **3**:448–456.
44. **Shani, M., U. Nudel, D. Zevin-Sonkin, R. Zakut, D. Givol, D. Katcoff, Y. Carmon, A. M. Frischauf, and D. Yaffe.** 1981. Skeletal muscle actin mRNA. Characterization of the 3′ untranslated region. Nucleic Acids Res. **9**:579–589.
45. **Shani, M., D. Zevin-Sonkin, Y. Carmon, H. Czosnek, U. Nudel, and D. Yaffe.** 1982. Changes in myosin and actin gene expression and DNase-I sensitivity associated with terminal differentiation of myogenic cultures, p. 189–200. *In* M. L. Pearson and H. F. Epstein (ed.), Muscle development. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
46. **Shih, C-K., M. Linial, M. M. Goodnow, and W. S. Hayward.** 1984. Nucleotide sequence 5′ of the chicken c-myc coding region: localization of a noncoding exon that is absent from myc transcripts in most avian leukosis virus-induced lymphomas. Proc. Natl. Acad. Sci. USA **81**:4697–4701.
47. **Solnick, D.** 1985. Trans splicing of mRNA precursors. Cell **42**:157–164.
48. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
49. **Steinmetz, M., A. Winoto, K. Minard, and L. Hood.** 1982. Clusters of genes encoding mouse transplantation antigens. Cell **28**:489–498.
50. **Tinoco, I. J., P. N. Borer, B. Dengler, M. D. Levine, O. C. Uhlenbeck, D. M. Crothers, and J. Gralla.** 1973. Improved estimation of secondary structure in ribonucleic acids. Nature (London) New Biol. **248**:40–41.
51. **Vandekerckhove, J., and K. Weber.** 1978. Mammalian cytoplasmic actins are the products of at least two genes and differ in primary structure in at least 25 identified positions from skeletal muscle actins. Proc. Natl. Acad. Sci. USA **75**:1106–1110.
52. **Vandekerckhove, J., and K. Weber.** 1979. The complete amino acid sequence of actins from bovine aorta, bovine heart, bovine fast skeletal muscle and rabbit slow skeletal muscle. Differentiation **14**:123–133.
53. **Vandekerckhove, J., and K. Weber.** 1984. Chordate muscle actins differ distinctly from invertebrate muscle actins. The evolution of the different vertebrate muscle actins. J. Mol. Biol. **179**:391–413.
54. **Zakut, R., M. Shani, D. Givol, S. Neuman, D. Yaffe, and U. Nudel.** 1982. Nucleotide sequence of the rat skeletal muscle actin gene. Nature (London) **298**:857–859.

CHAPTER 6

Mapping transcription start points on cloned genomic DNA with T4 DNA polymerase: a precise and convenient technique.

Mickey C-T. Hu and Norman Davidson

Gene 42:21-29 (1986)

GENE 1535

Mapping transcription start points on cloned genomic DNA with T4 DNA polymerase: a precise and convenient technique

(Recombinant DNA; single-stranded antisense DNA; 5′ end of mRNA; phage M13 vector; primer extension; strand displacement; S1 nuclease mapping)

Mickey Chien-Tsung Hu and Norman Davidson*

Department of Chemistry, California Institute of Technology, Pasadena, CA 91125 (U.S.A.) Tel. (818)356-6055

(Received June 21st, 1985)
(Revision received November 29th, 1985)
(Accepted December 2nd, 1985)

SUMMARY

We have developed a precise and convenient mapping technique for determining transcription start points (tsp) on cloned genomic DNA using T4 DNA polymerase. This method uses single-stranded (ss) M13 DNA and therefore is, unlike S1 and Exo VII nuclease mapping methods, independent of the restriction endonuclease sites present in the insert. Essentially the protocol involves the following steps: (1) hybridizing an mRNA to an ss M13 vector containing an antisense genomic DNA sequence spanning the presumptive tsp (cap site); (2) annealing a DNA primer (M13 sequencing primer) to the M13 DNA at a site on this DNA upstream from the 5′ end of the mRNA on the template DNA; (3) extending the DNA primer with T4 DNA polymerase towards the 5′ end of the mRNA. Since T4 DNA polymerase will not displace the mRNA : DNA hybrid, synthesis is blocked at the first nucleotide of the mRNA molecule. The length of the extended DNA products can then be determined with single nucleotide resolution on denaturing sequencing gels in parallel with a sequencing ladder. We have used this approach to map the tsp of the mouse skeletal α-actin gene. The sensitivity of the method allows precise mapping of transcripts present as 0.02–0.05% of the total RNA. This method is particularly valuable for mapping the tsp of genes which are known to contain a large intron between the first and second exons. It can also be applied to map the 5′ border of any given exon of a gene in an M13 vector or in other vectors that give ss DNAs.

* To whom correspondence and reprint requests should be addressed.

Abbreviations: BSA, bovine serum albumin; cDNA, DNA complementary to mRNA; dNTP, deoxynucleoside triphosphates; nt, nucleotide(s); PA, polyacrylamide; Pipes, piperazine-*N*,*N*′-bis[2-ethane sulfonic acid]; PolIk, Klenow (large) fragment of *E. coli* DNA polymerase I; RNasin, ribonuclease inhibitor; ss, single stranded; TBE, Tris–borate–EDTA; tsp, transcriptional start point(s); u, units.

INTRODUCTION

The study of individual genes from eukaryotes usually begins with the cloning of specific cDNA complementary to an mRNA or the cloning of a gene. The isolation and identification of genomic segments is then followed by transcriptional analysis to answer questions concerning development and gene regulation.

Two of the most popular techniques for mapping the 5′ ends of genes involve the use of ss-specific endonuclease S1 or exonuclease Exo VII to map RNA transcripts hybridized to well characterized template DNA (Berk and Sharp, 1978a). Both of these mapping techniques require precise restriction maps. If the labeled end of the DNA restriction fragment is within 200 nt of the end of the RNA and within the most 5′ exon, high-resolution mapping of the cap site can be obtained by S1 mapping. However, it has been found that S1 mapping can result in protected fragments several nt longer than corresponding to the cap site, due presumably to protection by the cap structure (Weaver and Weissmann, 1979). Exo VII data can be used to confirm the S1 mapping and are also useful for mapping short exons in which restriction sites have not been mapped (Berk and Sharp, 1978b). They are, however, not accurate for mapping the tsp of a gene which contains a large intron between the first and second exons, when the labeled site lies within the second exon.

Here, we describe an accurate and convenient procedure for mapping tsp on cloned genomic DNA with single nt resolution. This method uses ss M13 DNA and therefore is, unlike S1 or Exo VII mapping methods, independent of the restriction sites present in the insert. Furthermore, the sensitivity of the method makes possible precise mapping of moderately rare transcripts where sufficient RNA is available. It is particularly valuable for mapping the 5′ termini of genes which are known to contain a large intron between the first and second exons. It also provides an alternative to S1 and Exo VII mapping methods. We have successfully used this approach to map the tsp of the mouse skeletal α-actin gene. It should be possible to apply this method to map the 5′ border of any given exon of a gene cloned into an M13 vector or other vectors that give ss DNAs.

MATERIALS AND METHODS

(a) Materials

T4 DNA polymerase, T4 polynucleotide kinase, and PolIk were purchased from Bethesda Research Laboratories (BRL) or Boehringer-Mannheim. Radio-labeled nt were purchased from Amersham or New England Nuclear. Unlabeled nt were obtained from P-L Biochemicals, placental ribonuclease inhibitor, RNasin, from Promega Biotech, and the *Escherichia coli* 16S- and 23S-ribosomal RNAs from Boehringer-Mannheim.

(b) Preparation of DNA primer

The hexadecadeoxynucleotide (5′–GGGTAAC GCCAGGGTT–3′) was synthesized by S. Horvath, Caltech, on an automated DNA synthesizer (Hunkapiller et al., 1984). Purification of the oligodeoxynucleotide following synthesis was carried out using a 20% PA–8 M urea preparative gel in TBE buffer (90 mM Tris, 90 mM borate, 3 mM EDTA, pH 8.3). The desired product was eluted from the gel by a crush-soak method (Maxam and Gilbert, 1980) using 0.5 M NH_4 · acetate. An aliquot of the purified oligodeoxynucleotide was 5′ end labeled to a specific activity of 1.0×10^8 cpm/μg with polynucleotide kinase and [γ-^{32}P]ATP (approx. 3000 Ci/mmol) for use as a primer.

(c) Preparation of RNA and ss DNA

Total cellular RNA was isolated from differentiated cultures of the mouse myogenic cell line BC3H-1 (Schubert et al. 1974) using the guanidine thiocyanate-CsCl procedure (Chirgwin et al., 1979). Poly(A)$^+$ RNA was isolated from total cellular RNA using two cycles of oligo(dT)-cellulose chromatography (Aviv and Leder, 1982). The ss DNA was prepared basically as described by Sanger et al. (1980). All nucleic acids were purified by two cycles of precipitation with 2.5 vols. of ethanol and 0.1 vol. of 3 M Na · acetate (pH 5.5).

(d) RNA : DNA hybridization

All hybridizations and subsequent extension reactions were carried out in capped 0.5-ml eppendorf tubes. The dried RNA sample [10 or 5 μg of poly(A)$^+$ RNA] was dissolved in 2.2 μl diethylpyrocarbonate-treated H_2O, heated to 65 °C for 10 min and 2.8 μl of 9 × hybridization buffer (360 mM Na · Pipes, pH 6.4; 3.6 M NaCl; 9 mM EDTA) was added. One μl of M13 ss DNA (0.1 or 0.05 μg/μl)

with insert complementary to the 5′ region of the mRNA of interest and 20 μl of deionized formamide were added to the RNA solution. The mixture was denatured by heating at 80°C for 10 min and hybridized by incubation for 16 h at 37°C. Nucleic acids were precipitated by the addition of 35 μl of 5 M NH_4 · acetate and 250 μl of ethanol and chilled with dry ice for 30 min. The precipitate was collected by centrifugation at 4°C for 15 min, briefly dried, redissolved in 0.3 M Na · acetate and precipitated with 2.5 vols. of ethanol. The RNA : DNA pellets were washed with 75% ethanol and dried under vacuum.

(e) Primer extension with T4 DNA polymerase

Each of the dried RNA : DNA pellets was suspended in 9 μl of T4 DNA polymerase buffer [16.6 mM $(NH_4)_2SO_4$; 67 mM Tris · HCl, pH 8.8; 6.7 mM $MgCl_2$; 10 mM β-mercaptoethanol; 6.7 μM EDTA; 0.17 mg/ml BSA; 0.2 mM denatured salmon sperm DNA]. Note that the pellet should be dissolved directly in T4 DNA polymerase buffer rather than suspending in H_2O and adding 10 × T4 DNA polymerase buffer, subsequently. One μl (2.5 ng/μl) of the [^{32}P]DNA primer (1.0×10^8 cpm/μg) was added and the annealing mixture was heated to 65°C for 10 min and incubated for 1 h at 37°C. The synthesis reaction was carried out by adding 0.5 μl of 10 × T4 DNA polymerase buffer, 0.5 μl of RNasin (30 u/μl), 2 μl of 2.5 mM dNTP (a mixture of equal volume of 10 mM dATP; dGTP; dCTP; dTTP), and 1.5 μl of T4 DNA polymerase (5 u/μl, BRL). The reaction mixture was incubated for 1 h at 37°C. The reaction was then terminated by addition of EDTA to 10 mM and extraction with an equal volume of phenol–chloroform.

After ethanol precipitation and drying under vacuum, the RNA : DNA pellets were suspended in 2 μl of sterile distilled H_2O followed by 10 μl of gel loading buffer (90% formamide; 10 mM EDTA; 10 mM NaOH; 0.2% bromophenol blue; 0.2% xylene cyanol). The sample mixture was heated at 95°C for 3 min, quenched on ice, and electrophoresed on a 40 cm 6% PA–8 M urea sequencing gel in TBE buffer in parallel with four dideoxy DNA sequencing reactions as size markers. Following electrophoresis, the gel was immediately dried for 1 h on a BioRad gel dryer at 80°C and autoradiographed with Kodak XAR film at −70°C overnight.

(f) Primer extension with PolIk

The dried RNA : DNA pellets were dissolved in 10 μl of 50 mM NaCl, 7 mM Tris · HCl (pH 7.4), 10 mM $MgCl_2$, 3 mM dithiothreitol, containing 2.5 ng of the labeled DNA primer (1.1×10^7 cpm/μg) in a 0.5-ml eppendorf tube. The annealing mixture was heated to 65°C for 10 min and incubated at 45°C for 1 h, at which time an equal volume of the same annealing buffer, containing 1 mM each dNTP and 750 u/ml of RNasin, was added and the primer extended with PolIk (500 u/ml) for 1 h at an appropriate reaction temperature (14, 23, or 37°C). The reaction was terminated by addition of EDTA to 10 mM and the mixture was extracted with phenol–chloroform. After ethanol precipitation and drying under vacuum, the sample was heated at 95°C for 3 min in gel loading buffer (formamide/dye mix) and analyzed on a 40 cm 6% PA–8 M urea sequencing gel and autoradiographed.

(g) Nucleotide sequencing

The ss M13 templates were sequenced by the dideoxy chain termination procedure (Sanger et al., 1977) using [α-^{35}S]dATP (500 Ci/mmol) label (Biggin et al., 1983) with some modifications as described by Hu et al. (1986).

(h) Ribonuclease digestion

RNases A and T1 digestion of RNA : DNA hybrids to cleave off the overhanging ss RNA was performed as described by Melton et al. (1984).

RESULTS

(a) Mapping strategy

The basic principle of the method for mapping a tsp on a genomic clone is illustrated in Fig. 1. mRNA hybridized to the antisense strand of ss template DNA will serve as a "block" to extension by T4

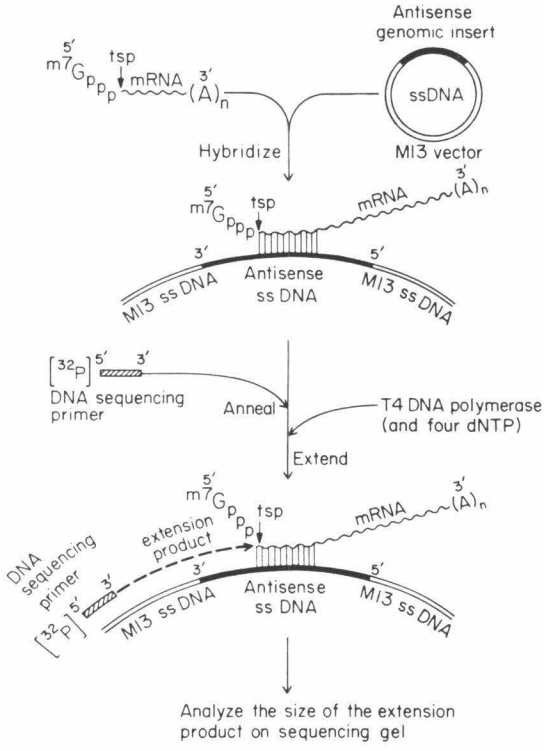


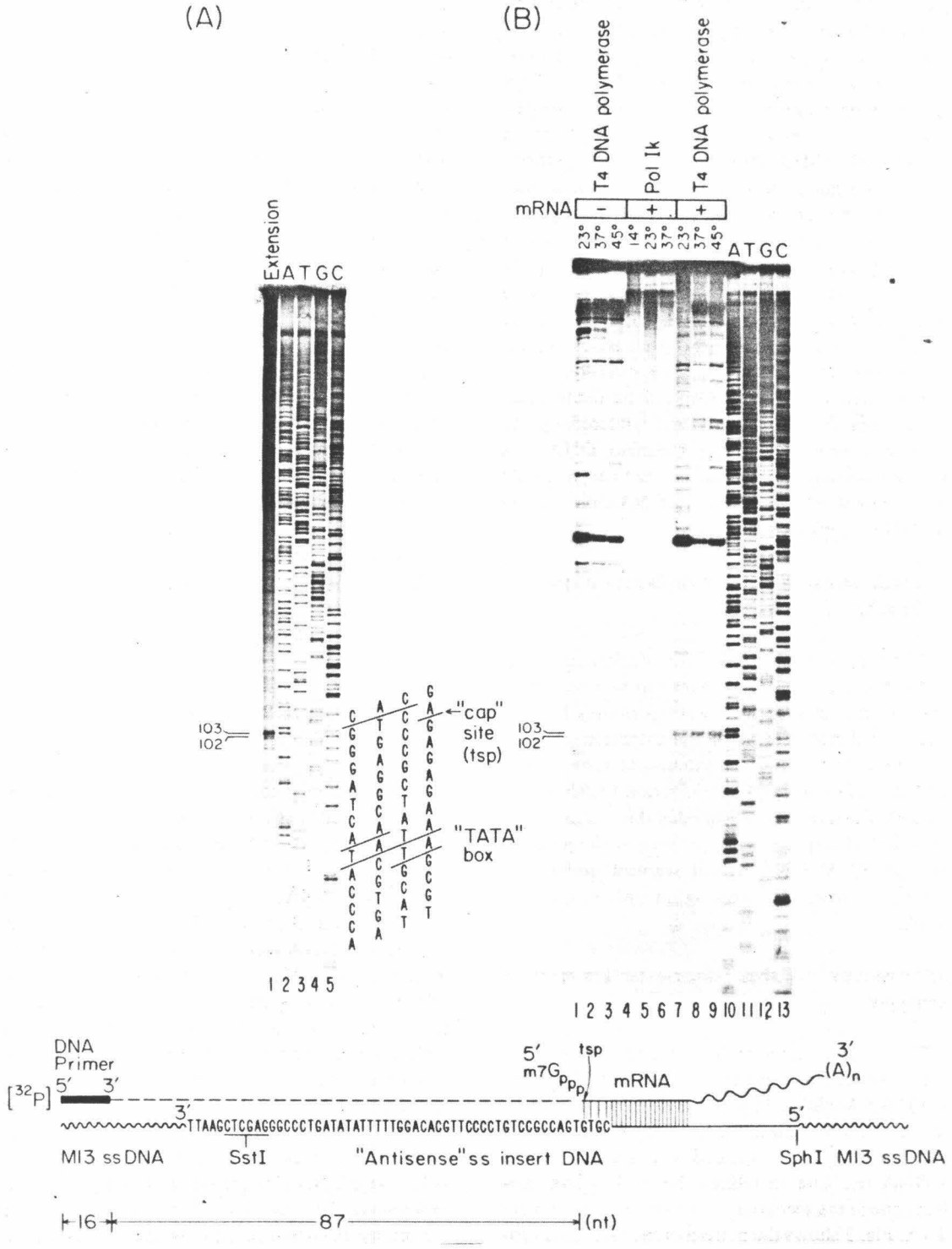
Fig. 1. Strategy for mapping transcription start points (tsp) on cloned genomic DNA.

DNA polymerase of a primer hybridized to the M13 ss DNA. Poly(A)$^+$ RNA is hybridized to the ss genomic sequence cloned in its antisense orientation into a phage M13 vector under conditions below the melting temperature (Tm) of the RNA : DNA hybrid (Casey and Davidson, 1977). Subsequently, a [^{32}P]DNA primer (M13 sequencing primer) is annealed to the M13 DNA, upstream from the 5′ end of the hybridized mRNA, and extended with T4 DNA polymerase. The synthesis is blocked by the 5′ end of the RNA since T4 DNA polymerase does not catalyze strand displacement synthesis (Masamune and Richardson, 1971). The length of the extended DNA products, determined with single nt resolution on a sequencing gel, can be compared with the result anticipated from the nt sequence.

(b) Mapping the tsp of the mouse skeletal α-actin gene

Fig. 2 shows the results of primer extension analyses of the tsp of the mouse skeletal α-actin gene. Poly(A)$^+$ RNA from differentiated BC3H-1 cells was hybridized to the ss α-actin genomic M13 clone containing the antisense DNA strand tentatively identified as containing the tsp of the gene by Exo VII mapping and the identification of the putative promoter region (Hu et al., 1986). In panel A, a major reaction product of 103 nt demonstrates the presence of a strong stop for the T4 DNA polymerase at the first "A" base of the nt sequence ACAC. Weaker bands of 102 and 101 nt are also evident. In the comparable experiment in Fig. 2B, the predominant band is again at 103 nt, with a distinct but weaker band at 102 nt. We suggest that the cap site corresponds to termination at 103 nt and that the mRNA begins with the sequence ACAC... In this interpretation, the band at 102 nt is due to premature termination of the primer extension product resulting from steric interference by the 5′ cap ($m^7G^{5'}ppp$). Alternatively, the mRNA actually begins with the sequence CACAC... The overall identification of the cap site is consistent with our lower resolution Exo VII mapping data (Hu et al., 1986)

Fig. 2. Primer-extension mapping of the tsp of the mouse skeletal α-actin gene. In the diagram in the lower part of the figure, the heavy line indicates the DNA primer and the dashed line indicates the bases added to the primer by T4 DNA polymerase; the lengths of the primer and of the anticipated extension (based on known sequence of the gene) are shown below. (A) Lane 1 [Extension], unlabeled DNA primer was hybridized to the template DNA in the presence of hybridized mRNA and extended with T4 DNA polymerase and [α-^{35}S]dATP (500 Ci/mmol) for 1 h at 37°C; lanes 2–5, sequencing ladders of the same template DNA in the absence of hybridized mRNA. The nucleotide sequence was shown on the right of the autoradiogram and the "cap" site and TATA box are underlined and indicated. (B) The ^{32}P-labeled primer was hybridized to: lanes 4–6, template DNA in the presence of hybridized mRNA and extended with PolIk; lanes 7–9, template DNA in the presence of hybridized mRNA and extended with T4 DNA polymerase. Controls: lanes 1–3, template DNA only and extended with T4 DNA polymerase; lanes 10–13, sequencing ladders. Reaction temperatures are indicated above the autoradiogram. The numbers on the left margins of the autoradiograms (A and B) denote the sizes (in nt) determined from marker sequence in adjacent lanes. All RNA : DNA hybridizations were performed with 3 μg of poly(A)$^+$ RNA hybridized to 1 μg of M13 ss template DNA.


(A)
(B)
Extension
A T G C
T4 DNA polymerase
Pol Ik
T4 DNA polymerase
mRNA
- + +
23° 37° 45° 14° 23° 37° 23° 37° 45°
A T G C
103
102
"cap" site (tsp)
"TATA" box
1 2 3 4 5
1 2 3 4 5 6 7 8 9 10 11 12 13
DNA Primer
5′ 3′
[32P]
5′ m7GpppA tsp mRNA 3′ (A)n
3′ TTAAGCTCGAGGGCCCTGATATATTTTTGGACACGTTCCCCTGTCCGCCAGTGTGC 5′
M13 ss DNA
SstI
"Antisense" ss insert DNA
SphI M13 ss DNA
16
87
(nt)

and as noted in the DISCUSSION, with the assignment for the rat α-actin mRNA (Zakut et al., 1982). The extension products (including the high-M_r smear on the autoradiogram) from the DNA without hybridization with RNA (Fig. 2B, lanes 1–3) suggest that there are several secondary structures existing in the ss template M13 DNA which can block synthesis by T4 DNA polymerase. This interpretation is supported by the increasing intensity of these bands when the extension reactions were performed at lower temperatures. We estimate that the template DNA, in this experiment, was in about fourfold excess over the actin mRNA, so many template molecules are not hybridized to that RNA. Therefore, it was not surprising that the intensity, on the autoradiogram, of the background products generated when the polymerase was hindered by the secondary structures in the template DNA was greater than that of the products obtained when the polymerase was blocked by the mRNA molecules on the DNA template.

(c) PolIk can catalyze strand displacement synthesis at the RNA : DNA hybrid

When primer extension was performed using PolIk (Fig. 2B, lanes 4–6), there was no discrete extension product even at low temperatures (14 °C). The overall background of the extension reaction was the same as that of the extension on the ss M13 template DNA without hybridization to RNA (not shown). These results indicate that this enzyme catalyzes strand-displacement synthesis at the junction with an RNA : DNA hybrid segment and at the secondary structures existing in the ss template DNA.

(d) Sensitivity of the primer-extension-mapping technique

To test the potential of the method, we carried out a reconstruction experiment using BC3H-1 poly(A)$^+$ RNA diluted with excess heterologous RNA, i.e., E. coli 16S- and 23S-ribosomal RNAs. In this experiment, we also titrated the amount of the ss DNA template to reduce the background products due to the secondary structures in the template DNA. Fig. 3 shows the pattern of the extension products from a serial dilution of poly(A)$^+$ RNA vs. 0.1 μg or 0.05 μg template DNA, and of 10 μg of total BC3H-1 RNA vs. 0.1 μg template DNA. It is clear that the intensity of the mRNA-specific signal (the bands labeled "cap site") increases corresponding to the amount of poly(A)$^+$ RNA used. As little as 0.1 μg of poly(A)$^+$ RNA, vs. 0.1 μg or 0.05 μg template DNA, can be detected with exposure time of 2 days without the aid of an intensifying screen using a [^{32}P]DNA primer with a specific activity of 1.0×10^8 cpm/μg DNA. Although skeletal α-actin mRNA is thought to be an abundant transcript, its actual abundance in the poly(A)$^+$ RNA fraction is not known. If the skeletal α-actin transcript is estimated to be present at 2% to 5% of the poly(A)$^+$ RNA from differentiated BC3H-1 cells, the detection of a band with 0.1 μg of BC3H-1 poly(A)$^+$ RNA diluted with 5 μg of heterologous RNA corresponds to the detection of an RNA of abundance 0.02 to 0.05% in a sample. This calculation suggests that the method is sufficiently sensitive to detect mRNA species present at levels of a few copies per cell (Affara et al., 1980). It is recommended, however, that poly(A)$^+$ RNA should be used rather than total RNA because of the higher background products in the latter case (as shown in lane 1).

(e) Overhanging 3′ RNA has no effect on the extension reactions

As an additional control, the following experiment was performed to ensure that the observed extension products were not artifacts of the interaction of the long overhanging 3′ portion of the bound RNA with the template DNA. The RNA : DNA hybridization was carried out as described above except that the antisense ss DNA was from an M13 subclone with a shorter insert which still contains the predicted tsp of the gene. Following hybridization, an aliquot of the mixture of the RNA : DNA hybrid was treated with RNases A and T1 to cleave off the overhanging RNA tail. The DNA primer was annealed to the ss template DNA of either the intact or the RNase-treated RNA : DNA hybrid, and then extended with T4 DNA polymerase. Primer extension on the ss template DNA alone was performed as a control experiment. The experimental rationale was that if the primer extension reaction of the T4 DNA polymerase was indeed blocked specifically by the 5′ end

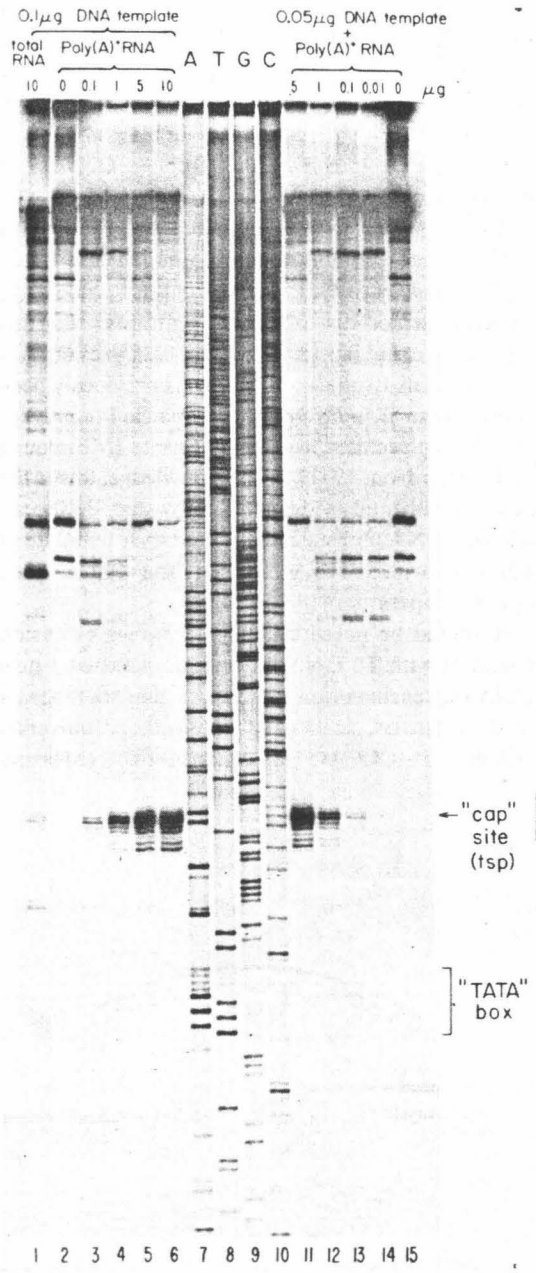


Fig. 3. Sensitivity test of the primer extension mapping technique with T4 DNA polymerase. Equal amounts of RNA samples (BC3H-1 poly(A)$^+$ RNA diluted with *E. coli* 16S and 23S ribosomal RNAs to make up a total of 5 μg of RNA) were hybridized to either 0.1 μg or 0.05 μg of M13 ss DNA containing the antisense strand of the mouse skeletal α-actin genomic clone DNA. The amount of poly(A)$^+$ RNA in each hybridization was as indicated above each lane (except that in lanes 1 and 6, 10 μg of total RNA and 10 μg of poly(A)$^+$ RNA were used). In each reaction, 2.5 ng of [^{32}P]DNA primer (1.0×10^8 cpm/μg) was hybridized to the template DNA and extended with T4 DNA polymerase for 1 h at 37°C. Following denaturation, equal amounts of radioactive counts of the primer extension samples were electrophoresed on a 6% PA–8 M urea gel. Lanes 7–10 demonstrate sequencing ladders of the same template DNA in the absence of hybridized mRNA. An autoradiographic exposure developed without the aid of an intensifying screen was used to map primer extension products. The "cap" site (tsp) and TATA box are indicated on the right of the autoradiogram.

of the mRNA on the template DNA, the predigested RNA molecule on the template DNA should have the same blocking effect on the synthesis of the polymerase as that of the intact RNA molecule on the template DNA. Furthermore, the extension products from different template DNAs, containing the same region of the RNA : DNA hybrid, should produce the same result. The primer extension results showed that the stop point for the extension reaction was not influenced by the RNA tail on the 3′ side of the exon, nor by slight changes in the position of the initiating primer relative to the tsp (not shown).

DISCUSSION

We describe here a new mapping technique for identifying tsp on cloned genomic DNA. The method utilizes the features of T4 DNA polymerase which allow it to carry out primed DNA synthesis on a ss template but prevent it from carrying out a strand displacement reaction through an RNA : DNA duplex region. Our measurements indicate an apparent variability of one nt in the stop position. The major band observed by us corresponds to a 5′ skeletal α-actin mRNA sequence of m^{7}G(5′)ppp(5′)ACAC. The weaker band with a length one nt less would correspond to the mRNA sequence of m^{7}G(5′)ppp(5′)CACAC. We suggest that the former assignment is correct because of the greater intensity of this band, but we do not have decisive proof of this point. If this interpretation is correct, the observed termination of the T4 DNA polymerase extension reaction to give a product one nucleotide shorter may be due to steric interference by the cap structure. We note that the conventional primer

extension and S1 mapping methods usually show several bands differing in length by one or more nt, with corresponding uncertainties in the cap site (Hanauer et al., 1983, Kost et al., 1983; Luse et al., 1981; Weaver and Weissmann, 1979).

The present method has one very significant advantage in addition to its apparent accuracy. In many cases, mouse skeletal α-actin being one of them, the first exon of a eukaryotic gene is quite short. It is then often difficult to find a useful restriction site within this exon for primer extension or S1 mapping. It is also frequently the case that the first intron is rather long, so that Exo VII mapping, using a restriction fragment labeled within the second exon, gives products which are too long to be measured with single nt accuracy.

The current method requires only that the insert, cloned in M13 or other vectors that give ss DNAs, spans the cap site. Indeed, it can be used to determine whether or not the cap site or a splice acceptor site of an exon (see below) is contained in a given insert. If the available insert extends too far upstream for accurate length measurements using a primer which lies within the vector sequence, standard methods for reducing the insert size can be used (Henikoff, 1984; Dale et al., 1985). Alternatively, a special primer within the insert could be synthesized and used.

In the conventional method of primer extension mapping of cap sites using reverse transcriptase and mRNA as template, secondary structures in the 5′ region of mRNA cause premature stops. Although secondary structures in the DNA sequence upstream from the cap site may cause similar problems in the present method, the lower thermal stability of DNA secondary structures compared to RNA secondary structures may provide some advantage for the present method.

The sensitivity of this method should also recommend it for a variety of studies. In the present application, we have mapped the cap site of an abundant transcript. Nevertheless, the sensitivity test experiment suggests that the experimental conditions used should work for a message present at only 0.02% to 0.05% of the poly(A)$^+$ RNA. Furthermore, we believe that by using optimal ratios of template DNA to poly(A)$^+$ RNA and by using a primer as close as possible to the presumptive cap site, thus minimizing the number of secondary structure stops, considerably greater sensitivity can be achieved. We further suggest that the method could be improved by adding T4 bacteriophage gene 32 protein in the primer extension reactions with T4 DNA polymerase to stimulate this polymerase by removing inhibitory secondary structure from the template DNA (Huberman et al., 1971).

T4 DNA polymerase is uniquely suited to this method since it contains neither endonuclease nor 5′ to 3′ exonuclease activity. Moreover, it does not catalyze strand-displacement synthesis. In this report, we demonstrate that PolIk catalyzes strand-displacement synthesis. While it has previously been shown that *E. coli* DNA polymerase I catalyzes strand-displacement synthesis at nicks (Masamune and Richardson, 1971), to our knowledge, this is the first experimental evidence to show that PolIk can displace RNA molecules from the DNA template at a junction with an RNA : DNA hybrid segment even at low temperature (14°C).

It should be noted that in the primer extension reactions with T4 DNA polymerase we used a new DNA sequencing primer as described in MATERIALS AND METHODS, section **b**. The synthetic universal primer 5′–CCCAGTCACGACGTT (Messing

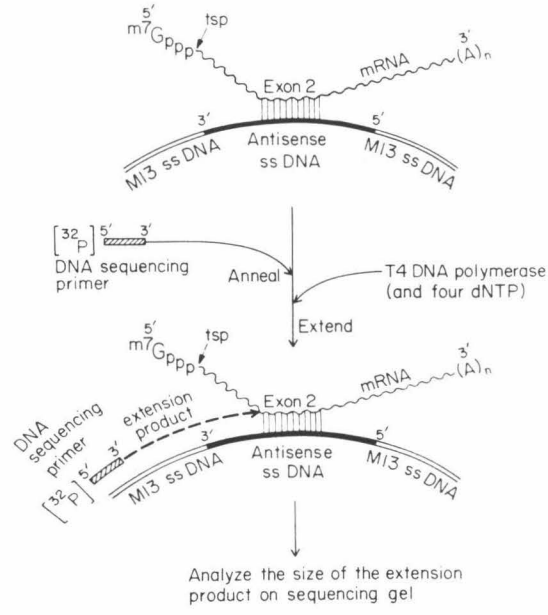


Fig. 4. Schematic representation of the proposed method for mapping the 5′ border of any given exon of a gene.

et al., 1981) has been found to hybridize to a second site on the M13 template, in the gene II region of the phage at position 787 (Messing, 1983). Alternatively, the new universal M13 sequencing primer 5′-AGTCACGACGTTGTA (Norrander et al., 1983) can be applied to this current method.

Finally, we note that if one has an M13 antisense insert spanning part of an exon and its upstream intron, the intron-exon border can be mapped. This proposed application is illustrated in Fig. 4.

ACKNOWLEDGEMENTS

We are grateful to Dr. Judith L. Campbell for technical advice and helpful comments on this manuscript. We also thank Drs. Bruce J. Nicholson, Sandra B. Sharp, Terry P. Snutch, and Michael M. White for their very valuable comments on the manuscript.

This work was supported by a research grant from the National Institutes of Health.

REFERENCES

Affara, N.A., Robert, B., Jacquet, M., Buckingham, M.E. and Gros, F.: Changes in gene expression during myogenic differentiation, I. Regulation of messenger RNA sequences expressed during myotube formation. J. Mol. Biol. 140 (1980) 441–458.

Aviv, H. and Leder, P.: Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid-cellulose. Proc. Natl. Acad. Sci. USA 69 (1972) 1408–1412.

Berk, A.J. and Sharp, P.A.: Spliced early mRNAs of simian virus 40. Proc. Natl. Acad. Sci. USA 75 (1978a) 1274–1278.

Berk, A.J. and Sharp, P.A.: Structure of the adenovirus 2 early mRNAs. Cell 14 (1978b) 695–711.

Biggin, M.D., Gibson, T.J. and Hong, G.F.: Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. Proc. Natl. Acad. Sci. USA 80 (1983) 3963–3965.

Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J.: Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry 18 (1979) 5294–5299.

Dale, R.M.K., McClure, B.A. and Houchins, J.P.: A rapid single-stranded cloning strategy for producing a sequential series of overlapping clones for use in DNA sequencing: application to sequencing the corn mitochondrial 18 S rDNA. Plasmid 13 (1985) 31–40.

Hanauer, A., Levin, M., Heilig, R., Daegelen, D., Kahn, A. and Mandel, J.L.: Isolation and characterization of cDNA clones for human skeletal muscle α actin. Nucl. Acids Res. 11 (1983) 3503–3516.

Henikoff, S.: Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene 28 (1984) 351–359.

Hu, M.C.-T., Sharp, S.B. and Davidson, N.: The complete sequence of the mouse skeletal α-actin gene reveals several conserved and inverted repeat sequences outside of the protein coding region. Mol. Cell. Biol. 6 (1986) 15–25.

Huberman, J.A., Kornberg, A. and Alberts, B.M.: Stimulation of T4 bacteriophage DNA polymerase by the protein product of T4 gene 32. J. Mol. Biol. 62 (1971) 39–52.

Hunkapiller, M., Kent, S., Caruthers, M., Dreyer, W., Firca, J., Giffin, C., Horvath, S., Hunkapiller, T., Tempst, P. and Hood, L.: A microchemical facility for the analysis and synthesis of genes and proteins. Nature 310 (1984) 105–111.

Kost, T.A., Theodorakis, N. and Hughes, S.H.: The nucleotide sequence of the chick cytoplasmic β-actin gene. Nucl. Acids Res. 11 (1983) 8287–8301.

Luse, D.S., Haynes, J.R., VanLeeuwen, D., Schon, E.A., Cleary, M.L., Shapiro, S.G., Lingrel, J.B. and Roeder, R.G.: Transcription of the β-like globin genes and pseudogenes of the goat in a cell-free system. Nucl. Acids Res. 9 (1981) 4339–4354.

Masamune, Y. and Richardson, C.C.: Strand displacement during deoxyribonucleic acid synthesis at single strand breaks. J. Biol. Chem. 246 (1971) 2692–2701.

Maxam, A.M. and Gilbert, W.: Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol., 65 (1980) 499–560.

Melton, D.A., Krieg, P.A., Rebagliati, M.R., Maniatis, T., Zinn, K. and Green, M.R.: Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. Nucl. Acids Res. 12 (1984) 7035–7056.

Messing, J.: New M13 vectors for cloning. Methods Enzymol., 101 (1983) 20–78.

Messing, J., Crea, R. and Seeburg, P.H.: A system for shotgun DNA sequencing. Nucl. Acids Res. 9 (1981) 309–321.

Norrander, J., Kempe, T. and Messing, J.: Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. Gene 26 (1983) 101–106.

Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. and Roe, B.A.: Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. J. Mol. Biol. 143 (1980) 161–178.

Sanger, F., Nicklen, S. and Coulson, A.R.: DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74 (1977) 5463–5467.

Schubert, D., Harris, A.J., Devine, C.E. and Heinemann, S.: Characterization of a unique muscle cell line. J. Cell Biol. 61 (1974) 398–413.

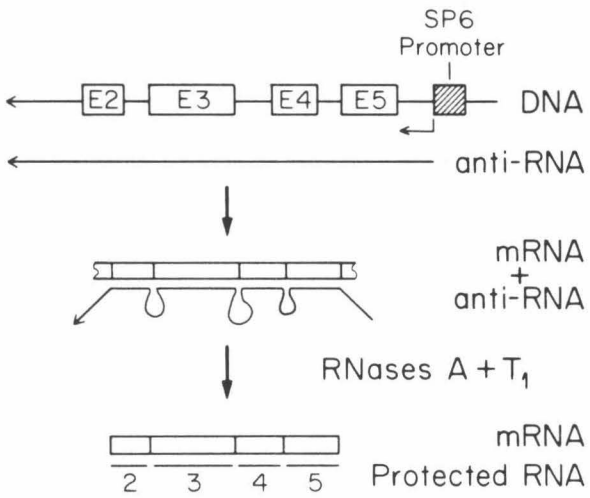
Weaver, R.F. and Weissmann, C.: Mapping of RNA by a modification of the Berk-Sharp procedure: the 5′ termini of 15S β-globin mRNA precursor and mature 10S β-globin mRNA have identical map coordinates. Nucl. Acids Res. 7 (1979) 1175–1193.

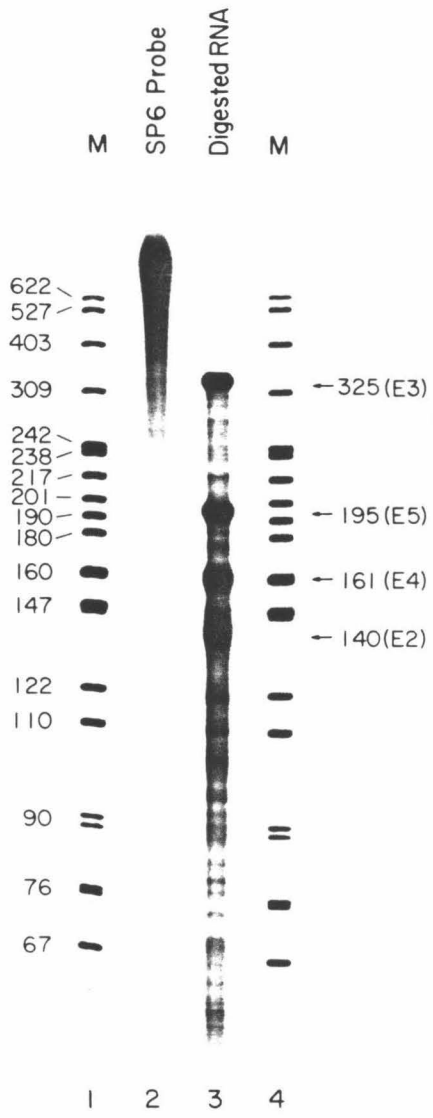
Zakut, R., Shani, M., Givol, D., Neuman, S., Yaffe, D. and Nudel, U.: Nucleotide sequence of the rat skeletal muscle actin gene. Nature 298 (1982) 857–859.

Communicated by S.R. Kushner.

APPENDIX A

Ribonuclease mapping of exons of the mouse skeletal α-actin gene with SP6 RNA probe.


SP6
Promoter
E2
E3
E4
E5
DNA
anti-RNA
mRNA
+
anti-RNA
RNases A + T1
mRNA
Protected RNA
2
3
4
5


M
SP6 Probe
Digested RNA
M
622
527
403
309
242
238
217
201
190
180
160
147
122
110
90
76
67
325(E3)
195(E5)
161(E4)
140(E2)
1
2
3
4

APPENDIX B

Codon usage in actin genes.

Codons \ Sources and genes	Human skeletal α	Mouse skeletal α	Rat skeletal α	Chick skeletal α	Rabbit skeletal α e	Human cardiac α	Chick cardiac α	Human nonmuscle β	Rat nonmuscle β	Chick nonmuscle β	Bovine nonmuscle*	Sea Urchin 1*	Sea Urchin 2	Sea Urchin SPG17	C.elegans I*	C. elegans II*	C. elegans III*	C. elegans IV*	Drosophila DmBF	Drosophila DmSC*	Drosophila DmI2A*	Drosophila DmS7A*	Drosophila DmB7E*	Drosophila Dm79B	Dictyostelium B1*	Dictyostelium 2(Sub1)*	Dictyostelium 2(Sub2)*	Dictyostelium 3*	Dictyostelium 5*	Dictyostelium MS*	Dictyostelium 7*	Acanthamoeba 1	Yeast	Soybean SAc1	Soybean SAc3	Maize MAc1	Codons	Average %	Vertebrates	Invertebrates	Plants
Ala GCA	1	0	0	0	0	2	5	2	1	2	0	0	1	1	0	0	0	0	1	1	4	0	2	2	0	0	0	0	0	0	0	0	0	11	9	1	Ala GCA	4.4	4.2	19.8
GCC	19	16	15	18	16	13	5	18	17	19	11	16	19	19	5	3	5	4	18	5	11	3	3	21	2	1	1	0	1	2	1	23	11	6	8	10	GCC	56.6	57.4	33.0
GCG	5	2	2	1	0	2	3	2	3	1	0	0	0	0	0	0	0	0	3	1	2	0	1	3	0	0	1	0	0	0	0	2	0	0	0	3	GCG	7.1	4.6	2.8
GCT	4	11	12	10	3	12	16	7	9	7	3	10	10	10	2	4	2	3	7	3	7	3	0	4	3	4	3	5	4	3	5	4	15	11	11	10	GCT	31.9	33.8	44.3
Arg AGA	1	0	0	1	0	1	1	1	1	6	0	1	2	2	2	2	2	2	0	0	0	0	0	0	1	1	1	1	1	1	2	0	13	2	6	6	Arg AGA	6.6	12.2	38.0
AGG	3	2	2	2	0	0	3	2	3	4	0	4	5	5	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	2	2	5	AGG	11.5	9.3	12.7
CGA	1	2	2	0	1	0	0	0	1	0	0	1	1	1	0	0	0	0	1	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	3	1	CGA	3.8	3.5	5.6
CGC	12	5	6	6	7	6	5	7	7	3	2	1	0	1	1	1	1	2	10	2	4	1	2	12	0	0	0	0	0	0	0	14	0	0	1	1	CGC	34.1	30.2	2.8
CGG	0	1	2	1	0	3	0	4	3	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	CGG	8.8	0.0	2.8
CGT	1	8	6	8	2	8	9	4	7	6	5	9	10	9	1	1	1	0	6	4	8	2	3	6	2	2	2	2	2	2	2	3	5	12	6	4	CGT	35.2	44.8	38.0
Asn AAC	10	9	9	9	8	5	7	7	6	6	5	9	9	9	1	1	1	1	8	1	7	0	1	8	0	1	0	0	1	0	0	9	9	6	4	8	Asn AAC	72.3	81.7	79.4
AAT	2	3	3	3	0	7	5	2	3	3	0	0	0	0	1	1	1	1	1	0	2	1	1	1	1	0	1	1	0	1	1	0	0	2	4	1	AAT	27.7	18.3	20.6
Asp GAC	18	18	18	16	7	13	8	18	18	8	10	13	15	15	7	9	7	8	16	5	12	4	5	16	1	0	0	1	1	1	2	18	9	8	6	7	Asp GAC	69.7	63.9	34.1
GAT	4	4	4	6	0	9	14	5	5	15	0	8	8	8	2	1	2	2	8	2	4	2	4	4	4	5	4	3	5	4	5	3	11	14	17	16	GAT	30.3	36.1	65.9
Cys TGC	5	5	4	5	3	3	4	4	4	4	2	3	6	5	1	1	1	1	6	2	3	1	2	6	0	0	0	0	0	0	0	4	0	2	3	3	Cys TGC	72.9	65.6	42.1
TGT	1	1	2	1	0	3	2	2	2	2	0	1	1	1	1	1	1	1	1	1	3	1	1	1	1	1	1	1	1	1	1	0	4	4	1	2	TGT	27.1	34.4	57.9
Gln CAA	0	1	1	4	0	3	3	1	1	0	0	2	1	1	2	2	2	1	1	1	5	1	0	3	2	2	2	2	2	2	3	1	14	6	7	2	Gln CAA	12.0	32.5	67.4
CAG	11	10	10	7	9	8	8	11	11	12	6	7	12	12	1	1	1	2	10	3	6	1	3	10	0	0	0	0	0	0	0	10	0	5	2	7	CAG	88.0	67.5	32.6
Glu GAA	3	3	3	7	1	7	16	2	5	7	2	3	6	6	0	0	0	1	0	2	6	1	0	4	1	1	1	2	1	1	2	0	26	8	12	8	Glu GAA	19.6	18.3	49.5
GAG	25	25	25	21	22	21	12	24	21	19	14	18	20	20	4	2	4	2	26	3	15	0	4	24	0	0	0	0	0	0	0	28	1	17	17	20	GAG	80.4	81.7	50.5
Gly GGA	0	0	0	1	0	2	2	0	1	2	0	9	13	13	8	8	8	8	1	3	5	0	4	2	0	0	1	0	0	1	0	1	0	12	8	4	Gly GGA	3.0	25.8	20.0
GGC	20	13	13	12	7	15	7	23	15	8	4	4	4	5	0	0	0	0	12	6	13	3	3	17	0	0	0	0	0	0	0	16	0	4	4	8	GGC	50.9	25.7	13.3
GGG	4	5	5	5	2	3	2	3	2	1	3	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	3	3	6	GGG	13.0	0.3	10.0
GGT	4	10	10	10	1	8	17	2	9	17	1	11	11	10	3	3	3	3	16	5	5	5	4	10	9	9	4	8	9	8	11	13	29	13	14	12	GGT	33.1	48.6	56.7
His CAC	9	7	8	7	5	6	6	8	7	6	6	6	7	7	1	0	1	1	7	3	6	1	2	9	1	0	0	0	1	0	0	10	7	2	3	6	His CAC	80.6	77.8	46.2
CAT	0	2	1	2	1	3	3	1	2	3	0	2	2	2	1	2	1	1	2	0	1	0	1	1	0	1	0	0	0	0	1	0	3	7	7	4	CAT	19.4	22.2	53.8
Ile ATA	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	2	1	1	Ile ATA	0.3	0.4	3.7
ATC	28	27	25	22	20	18	15	24	20	18	9	16	26	26	3	2	3	4	24	3	14	1	6	25	1	0	1	0	1	0	1	24	16	7	11	11	ATC	76.1	80.8	42.1
ATT	2	3	5	8	1	12	14	4	8	10	3	1	1	2	2	2	2	1	5	3	7	0	1	2	1	2	0	2	1	2	2	3	14	16	16	12	ATT	23.6	18.8	54.2
Leu CTA	0	0	0	1	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	3	0	0	0	0	0	0	2	0	2	1	3	2	Leu CTA	2.2	2.8	7.5
CTC	6	5	5	5	4	6	5	6	6	7	3	11	12	11	1	1	1	1	2	1	2	0	1	3	0	0	0	0	0	0	0	18	0	4	9	8	CTC	21.4	30.4	19.6
CTG	18	19	17	16	11	16	10	20	19	18	11	6	4	4	0	0	0	0	24	2	5	1	1	18	0	0	0	0	0	0	0	8	0	8	0	5	CTG	64.6	34.1	12.1
CTT	2	1	2	1	0	0	3	0	0	0	1	6	10	10	2	2	2	2	0	1	4	0	0	2	0	0	0	1	0	1	0	3	2	9	11	6	CTT	3.7	21.5	26.2
TTA	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	2	0	1	0	0	0	1	1	1	1	1	1	1	0	2	1	0	0	TTA	0.7	4.7	2.8
TTG	0	1	2	3	0	3	7	0	1	1	2	1	1	2	0	0	0	0	1	0	6	0	1	2	0	0	0	0	0	0	0	0	19	6	4	5	TTG	7.4	6.5	31.8
Lys AAA	5	5	5	3	1	3	8	3	3	5	1	1	1	1	0	2	0	0	2	0	2	0	3	0	2	2	2	2	2	2	3	0	6	4	7	2	Lys AAA	22.1	15.9	25.0
AAG	14	14	14	16	11	16	11	16	16	13	7	11	17	18	5	2	5	5	17	6	14	2	3	19	0	0	0	0	0	0	0	19	12	15	13	17	AAG	77.9	84.1	75.0
Met ATG	17	17	17	17	9	16	16	17	17	17	8	16	17	16	5	5	5	5	15	6	14	4	4	16	4	4	1	4	4	4	4	16	16	15	17	18	Met ATG	100.0	100.0	100.0
Phe TTC	11	9	9	8	10	8	7	12	12	12	5	10	13	13	2	2	2	2	12	2	7	2	3	11	1	1	2	1	1	1	1	11	12	6	6	7	Phe TTC	80.5	84.7	66.0
TTT	1	3	3	4	0	4	5	1	1	1	2	1	1	1	0	0	0	0	1	1	3	0	0	1	1	1	1	1	1	1	1	2	2	5	5	4	TTT	19.5	15.3	34.0
Pro CCA	1	2	2	4	0	1	4	1	2	4	0	6	10	10	4	3	4	4	4	1	2	1	0	3	3	3	2	2	3	2	3	0	13	8	7	6	Pro CCA	11.0	38.9	43.6
CCC	9	9	10	10	11	9	6	12	6	9	6	8	9	9	0	0	0	0	11	4	7	2	4	12	0	0	0	0	0	0	0	15	0	1	6	3	CCC	50.8	45.0	12.8
CCG	5	2	2	0	1	3	1	0	0	0	0	0	0	0	0	0	0	0	1	1	6	0	0	4	0	0	0	0	0	0	0	3	0	2	0	2	CCG	7.3	8.3	5.1
CCT	4	6	5	5	0	6	8	6	11	6	2	1	1	1	0	1	0	0	3	1	2	0	1	0	0	0	0	1	0	1	0	1	6	9	5	10	CCT	30.9	7.8	38.5
Ser AGC	4	4	4	4	3	5	5	6	5	7	2	3	5	4	0	0	0	0	1	2	2	0	1	2	0	0	0	0	0	0	0	0	0	8	7	11	Ser AGC	20.3	9.4	22.2
AGT	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1	0	0	0	0	2	2	7	3	AGT	0.0	1.9	12.0
TCA	0	2	2	0	0	0	2	0	1	0	0	4	3	3	0	0	0	0	1	0	2	1	0	1	0	1	0	0	1	0	2	0	3	7	5	2	TCA	2.9	8.9	14.5
TCC	15	12	11	16	11	11	10	16	14	9	3	7	8	13	3	3	3	4	11	1	8	1	2	16	0	0	0	0	0	0	0	6	12	4	5	6	TCC	53.1	40.4	23.1
TCG	4	1	1	0	2	0	2	1	0	0	0	0	1	0	1	1	1	0	11	3	7	1	1	4	0	0	0	0	0	0	0	17	0	0	0	3	TCG	4.6	22.5	2.6
TCT	0	4	5	3	1	8	5	2	5	9	4	5	6	5	0	0	0	0	0	1	5	0	0	2	2	1	1	2	1	2	2	1	14	6	7	3	TCT	19.1	16.9	25.6
Thr ACA	1	2	2	7	1	5	9	2	4	9	2	0	3	2	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	7	3	4	Thr ACA	16.2	3.4	16.9
ACC	22	20	20	16	12	12	13	17	15	12	8	13	16	16	2	2	2	2	22	4	9	0	2	20	0	0	0	0	0	0	1	22	8	8	6	7	ACC	61.4	75.6	34.9
ACG	3	1	1	0	3	0	0	2	0	1	4	0	0	1	0	0	0	0	3	0	2	0	0	1	0	0	0	0	0	0	0	1	0	2	0	5	ACG	5.5	4.5	8.4
ACT	1	4	4	4	2	9	4	5	7	4	2	4	3	3	0	0	0	0	1	0	7	0	0	2	1	1	0	1	1	1	1	3	13	5	9	6	ACT	16.9	16.5	39.8
Trp TGG	4	4	4	4	4	4	4	4	4	4	3	2	3	4	1	1	1	1	4	0	4	0	2	4	0	0	0	0	0	0	0	4	4	4	4	4	Trp TGG	100.0	100.0	100.0
Tyr TAC	14	9	10	8	9	10	5	12	10	10	7	14	14	15	2	2	2	2	12	3	9	0	1	10	0	0	0	0	0	0	0	14	14	5	2	10	Tyr TAC	65.4	85.5	55.4
TAT	2	7	6	8	0	6	11	3	5	5	2	0	0	0	0	0	0	0	3	0	2	0	1	6	0	0	1	1	0	1	1	1	0	9	11	5	TAT	34.6	14.5	44.6
Val GTA	0	1	0	1	0	2	1	0	2	2	0	0	1	1	1	1	1	1	0	2	0	2	0	1	4	0	0	1	0	0	0	0	0	0	3	3	3	Val GTA	4.2	5.9	7.8
GTC	6	4	5	8	0	10	8	8	5	5	4	10	10	10	5	5	5	4	5	3	7	4	3	8	1	1	0	0	1	0	0	15	12	4	7	12	GTC	29.7	37.9	30.2
GTG	15	16	16	12	10	6	8	13	13	10	6	3	3	3	1	0	1	3	13	4	6	1	3	10	0	0	0	0	0	0	0	6	0	8	5	5	GTG	59.0	22.3	15.5
GTT	0	0	0	0	0	3	4	1	2	5	0	9	9	9	2	3	2	2	1	2	6	2	2	1	5	5	3	5	5	5	7	2	14	16	13	11	GTT	7.1	34.0	46.6
Total codons analyzed	377	377	377	377	232	377	377	375	375	375	183	305	376	376	86	83	86	86	376	108	308	53	93	376	51	51	41	51	52	51	68	375	375	375	375	375					

PROPOSITIONS

PROPOSITION 1

Expression cloning of cDNAs encoding the p75 subunit of the high-affinity interleukin-2 receptor by a new selection strategy.

ABSTRACT

Interleukin-2 (IL-2) binds to both high- and low-affinity IL-2 receptors on activated T lymphocytes. A critical question concerning the role of IL-2 receptors (IL-2-R) in the control of T-cell growth, is the mechanisms by which IL-2 bound to high-affinity membrane receptors can be internalized and transduce a growth signal to the cell nucleus. Current data are consistent with the view that high-affinity IL-2 receptors are composed of p75 and p55 heterodimeric receptors, a p75 subunit noncovalently linked to a p55 subunit, each of which is independently capable of binding IL-2 with much lower affinity. The IL-2-R p75 subunit appears to be the biologically relevant moiety, whereas the p55 subunit seems to function merely as a helper binding protein, having no signaling capacity of its own. The IL-2-R p55 subunit has been extensively studied, and its cDNA clones have been isolated. However, the IL-2-R p75 subunit has not been purified, nor have its cDNA clones been isolated. I therefore propose two major steps to directly isolate cDNA clones encoding the IL-2-R p75 subunit without prior purification of this protein. First, I will construct a selected cDNA library by using subtraction probes to enrich for IL-2-R p75 cDNAs. Second, IL-2-R p75-specific cDNA clones will be isolated from this library using a new selection method- an indirect immunoselection protocol. If successful, this should contribute to an understanding of the molecular basis of signal transduction by the high-affinity IL-2 receptor. Moreover, the proposed selection strategy may represent a powerful general approach for the isolation of genes encoding receptors for other growth factors.

INTRODUCTION

The activation of a human T cell immune response to a foreign antigen is initiated by the interaction of the human T cell antigen receptor (for a review, see Allison and Lanier, 1987) with antigen presented on the cell surface in the context of products of the major histocompatibility locus. This activation also requires the macrophage-derived interleukin-1 (for a review, see Oppenheim et al., 1986). Following this interaction, a subset of helper T cells express the gene encoding the T-cell growth factor interleukin-2 (IL-2) (Smith, 1980). To exert its biological effect, IL-2 must interact with specific high-affinity membrane receptors. Activated lymphocytes with high-affinity receptors can proliferate in response to IL-2, resulting in the expansion of T cells with specialized effector functions, including helper, suppressor, and cytotoxic activities. Elucidation of the mechanism of the interaction of IL-2 with IL-2 receptors is essential to understanding the dynamics of IL-2 mediated T-cell growth and for the exploitation of its therapeutic potential. A failure of the production of either the growth factor IL-2, or its receptor, results in failure of the T cell immune response. Although the interaction of an appropriately presented antigen with its specific polymorphic receptor complex confers specificity for a given T-cell immune response, the interaction of IL-2 with IL-2 receptors determines its magnitude and duration.

Activated T cells express two classes of IL-2 receptor (IL-2-R) that differ in their affinity for IL-2. In both normal activated lymphocytes and most IL-2-R expressing T-cell lines tested, only 1-10% of the receptors are high affinity (dissociation constant $K_d \approx 10^{-11}$M)

and appear to mediate the physiological response of T cells to IL-2; the remaining receptors bind IL-2 with much lower affinity ($K_d \approx 10^{-8}$M) (Robb et al., 1981; Robb et al., 1984; Lowenthal et al., 1985). Current evidence suggests that only IL-2 bound to high-affinity receptors is internalized via receptor-mediated endocytosis (Weissman et al., 1986; Fujii et al., 1986). Both classes of receptors share the same Tac peptide (the IL-2-R p55 subunit) defined by the anti-Tac monoclonal antibody (Robb et al., 1984; Waldmann, 1986). The IL-2-R p55 subunit has been extensively studied, its cDNA clones have been isolated, and the structure of the gene determined (Leonard et al., 1984; Nikaido et al., 1984; Cosman et al., 1984; Leonard et al., 1985; Ishida et al., 1985). However, the molecular basis for the distinction in affinity between high- and low-affinity receptors has not been elucidated. Recent reports from several laboratories provide strong evidence that there is a second IL-2 binding protein (p75 subunit) that interacts with the IL-2-R p55 subunit to form the high affinity IL-2 receptor (Sharon et al., 1986; Tsudo et al., 1986; Teshigawara et al., 1987; Tsudo et al., 1987; Robb et al., 1987; Dukovich et al., 1987).

Several observations challenged the idea that the IL-2-R p55 subunit was the only molecule involved in binding IL-2 and triggering a physiological response. First, transfection of nonlymphoid mouse cells with cDNAs encoding human IL-2-R p55 generated only low-affinity receptors (Greene et al., 1985) whereas similar transfection of two mouse T-cell lines displayed both high- and low-affinity human IL-2 receptors (Hatakeyama et al., 1985; Kondo et al., 1986). These results suggested that T cells express a cell-specific component that either

directly contributes to the binding site of the high-affinity IL-2 receptor or modulates the binding affinity of IL-2-R p55. Second, the deduced structure of the IL-2-R p55 predicts that only 13 amino acids form the cytoplasmic domain, raising questions as to the mechanism of signal transduction (Leonard et al., 1984; Nikaido et al., 1984; Cosman et al., 1984). In fact, IL-2 saturation of the low affinity binding sites expressed by normal T cells has no effect on the rate of cell proliferation, and non-T cells transfected solely with the IL-2-R p55 cDNA can not respond to IL-2 (Smith, 1988). Third, some lymphocytes, including natural killer cells and precursors of lymphokine-activated killer cells, apparently do not express the IL-2-R p55 but still bind and respond to IL-2.

The available data are consistent with the view that high-affinity IL-2 receptors consist of at least two subunits, each of which is independently capable of binding IL-2 with much lower affinity. It has been shown that each subunit interacts with a different region of the IL-2 peptide molecule, supporting the idea that both subunits directly contribute to the binding site of the high-affinity receptor (Robb et al., 1987). Recently, it has been demonstrated that IL-2 binds to and dissociates from the IL-2-R p75 subunit 10^2 to 10^3 times more slowly than it does to the p55 subunit, which takes place very rapidly (Wang and Smith, 1987). Furthermore, it has been indicated that the IL-2-R p75 subunit by itself is sufficient to signal T-cell proliferation (Sharon et al., 1986), suggesting that the IL-2-R p75 subunit is the biologically relevant moiety whereas the p55 subunit appears to function merely as a helper binding protein, having no signaling

capacity of its own. Nevertheless, the generation of an extremely high affinity receptor from two distinct low affinity binding proteins may be of even greater importance for regulation of IL-2 responses.

The notion that high-affinity IL-2 receptors are p75/p55 heterodimeric receptors has important implications. First, structural studies of the p75 subunit may help to elucidate signal transduction by the high-affinity IL-2 receptor to promote T-cell growth. Second, functional studies of this gene (p75 subunit) expression in normal T cells may help to explain the role of IL-2 during thymic differentiation, since regulated IL-2-R expression is universally associated with mature T-cell immune competence. Therefore, understanding the mechanism of the interaction of IL-2 with its high-affinity receptors (p75/p55) is of fundamental importance in developing an overall picture of the operation of the T cell immune system.

SPECIFIC AIMS

The primary objectives of this proposal are to isolate cDNA(s) encoding the p75 subunit of the high-affinity IL-2 receptor, investigate the structure of the peptide deduced from them, and study the regulation of this gene (p75) expression in normal human T cells. The research project proposed here represents a promising approach in which to study how the IL-2-R p75 subunit interacts with the p55 subunit to form the high-affinity IL-2 receptor and mediate the physiological response of T cells to IL-2. Ultimately, the results of this project will provide evidence as to how IL-2 specifically functions on T cells. This should lead, in the long term, to an understanding of the

molecular basis of signal transduction by the high-affinity IL-2 receptor and thus contribute to the elucidation of the mechanisms governing the control of thymic differentiation and the development of the functional capacities of the T lymphocyte.

METHODS AND PROCEDURES TO BE USED IN PROPOSED PROJECT

I propose two major steps to directly isolate cDNA clones encoding the IL-2-R p75 subunit without prior purification of this protein. First, I will construct a selected cDNA library by using subtraction probes to enrich for IL-2-R p75 cDNAs as described (Hedrick et al., 1984). Second, IL-2-R p75 subunit-specific cDNA clones will be isolated from this library by using a new method - an indirect immunoselection protocol (see below). Overall, this cloning approach is based on the following assumptions about the nature of the IL-2-R p75 gene(s): (1) That they should be expressed in YT-clone-2C2 cells, a human leukemic T cell clone that expresses solely the IL-2-R p75 (Teshigawara et al., 1987), but not in MT-1 cells, a human leukemic T cell line that does not express the IL-2-R p75 subunit (Tsudo et al., 1987). (2) That the mRNAs for the IL-2 receptor proteins should be found on membrane bound polysomes, as one would expect the nascent receptor polypeptides to attach to the endoplasmic reticulum by a leader peptide (Blobel and Dobberstein, 1975). Therefore, by synthesizing ^{32}P-labeled cDNAs of the membrane-bound polysomal RNA of YT-clone-2C2 cells and subtracting with those sequences also expressed in MT-1 cells, one should be left with a specific probe representing a small fraction of total YT-clone-2C2 cell gene expression and likely to include copies of the mRNAs that encode

the IL-2-R p75 subunit, but not the p55 subunit. (3) That the IL-2-R p75 subunit itself is capable of binding IL-2 specifically (Teshigawara et al., 1987; Tsudo et al., 1987a and b). (4) That it is possible to obtain a non-blocking anti-IL-2 monoclonal antibody or a polyclonal antiserum to IL-2 (see below).

The strategy to be used to isolate cDNA clones representing membrane-bound mRNAs expressed in YT-clone-2C2, but not in MT-1 cells, is shown in Figure 1A (page 91). Membrane-bound polysomal poly(A)$^+$ RNA will be prepared from YT-clone-2C2 cells as described (LaPolla et al., 1984); total cellular poly(A)$^+$ RNAs from YT-clone-2C2 and MT-1 cells will be prepared by the procedure as described (Chirgwin et al., 1979). (These two cell lines can be obtained from Dr. Kendall A. Smith, Dartmouth Medical School, Hanover, NH.) ^{32}P-labeled cDNA and unlabeled cDNA will be synthesized from membrane-bound and total cellular poly(A)$^+$ RNAs, respectively, using oligo(dT) and reverse transcriptase (Davis et al., 1984). YT-clone-2C2 cDNA synthesized from total cellular poly(A)$^+$ RNA will be subtracted with MT-1 total poly(A)$^+$ RNA and the subtraction products will be used to synthesize the second strand DNA with DNA polymerase I as described (Davis et al., 1984). [The separation of cDNA from RNA and from DNA:RNA hybrids will be carried out by using a CsCl gradient containing guanidine hydrochloride as described (Lubbert et al., 1987).] The double-stranded cDNA will then be cloned into the πH3M vector as described (Seed and Aruffo, 1987), and the ligated DNA will be transformed into *E. coli* competent cells to generate a plasmid cDNA expression library. ^{32}P-labeled cDNA probes of YT-clone-2C2 cells will be subtracted with MT-1 total cellular poly(A)$^+$

RNA, and used to screen the above cDNA library. After being screened with the YT-clone-2C2-specific probe, the selected cDNA library can be rescreened.

The strategy of the indirect immunoselection protocol is depicted in Figure 1B (page 91). Specifically, in this protocol a spheroplast fusion technique (Sandri-Goldin et al., 1981) will be used to introduce the library into COS cells, where it replicates and expresses its inserts. The cells in dishes will be treated with recombinant human IL-2 from *E. coli* (Wang et al., 1984), which is commercially available, at a final concentration of $5x10^{-8}M$ and incubated for 1 hour at 4 ^{o}C according to the method described by Tsudo et al. (1986). It is crucial here to assume that the IL-2-R p75 subunit expressed by itself in COS cells will bind IL-2 since it has been demonstrated that in some T cells the p75 subunit alone, without the p55 subunit, can bind IL-2 specifically (Teshigawara et al., 1987; Tsudo et al., 1987a and b). After washing off the unbound IL-2, the cells will be harvested by detaching without trypsin, treated with monoclonal antibody specific for human IL-2 (Smith et al., 1983), and then distributed in dishes coated with affinity-purified antibody to mouse immunoglobulins as described (Seed and Aruffo, 1987). It should be noted that the anti-IL-2 monoclonal antibody to be used here should be a non-blocking antibody (i.e. this antibody recognizes a site of IL-2 peptide that will not interfere with binding of IL-2 to its receptors). Alternatively, a polyclonal antiserum to IL-2 could be used. Probably, some of the antibodies therein will recognize epitopes on the surface of IL-2 that are not involved in binding to the receptors. (Some mouse monoclonal

antibodies and polyclonal antisera to human IL-2 may be obtained from Dr. Kendall A. Smith, Dartmouth Medical School, Hanover, NH.) The rationale of this indirect immunoselection protocol is that cells expressing IL-2-R p75 subunit will be bound with IL-2 which in turn will be recognized by the non-blocking anti-IL-2 monoclonal antibody or an anti-IL-2 antiserum so that these cells will adhere to dishes (coated with anti-mouse IgG antibody), and the remaining cells can be washed away. The adherent cells should express solely the IL-2-R p75 but not the p55 subunit since the p55 subunit sequences have been depleted from the cDNA expression library as described above. Further, this can be verified by treating those cells with the anti-p55 (anti-Tac) monoclonal antibody before performing IL-2 binding assay (Robb et al., 1981; Tsudo et al., 1986), and they should not be blocked. From the adherent cells, plasmid DNAs can be recovered by the "Hirt" procedure (Hirt, 1967) and transformed back into *E. coli* for further rounds of transfection and selection. Whether or not the positive cDNAs are derived from IL-2-R p75 mRNA will be determined by transfecting these plasmid DNAs into MT-1 cells, which constitutively express p55 but not p75 subunit (Tsudo et al., 1987), to reconstitute high-affinity IL-2 receptors. Moreover, the IL-2-R p75-producing cells can be confirmed by an IL-2 cross-linking study, as described (Tsudo et al., 1986; Teshigawara et al., 1987), to see whether they have expressed the p75 peptide. It is important to reiterate that the strategy described above using the πH3M vector (Aruffo and Seed, 1987) for the expression cDNA library is chosen because of the following reasons: (1) That the eukaryotic transcription unit allows high-level expression in COS cells

of coding sequences placed under its control. (2) That the small size and particular arrangement of sequences in this vector permit high-level replication as a plasmid in COS cells.

The strategy described above using an indirect immunoselection procedure to isolate cDNA clones for the IL-2-R p75 subunit may prove useful in the isolation of clones for other membrane proteins. Specifically, in cases where it is difficult to purify membrane proteins, due to their low abundance and their hydrophobic nature, and/or to make antibodies, this approach requires only a second non-blocking antibody or a polyclonal antiserum reactive to the ligand which can directly bind to the specific membrane protein. For instance, this selection strategy should be applicable for the isolation of genes encoding receptors for other growth factors such as the human hematopoietic colony-stimulating factors, including granulocyte colony-stimulating factor (G-CSF), macrophage colony-stimulating factor (M-CSF), granulocyte-macrophage colony-stimulating factor (GM-CSF), and interleukin-3 (IL-3).

REFERENCES

1. Allison, J.P., and Lanier, L.L. (1987) *Ann. Rev. Immunol.* 5, 503-540.

2. Blobel, G., and Dobberstein, B. (1975) *J. Cell Biol.* 67, 835-851.

3. Chirgwin, J.M., Przybyla, A.E., MacKonald, R.J., and Rutter, W.J. (1979) *Biochemistry* 18, 5294-5299.

4. Cosman, D., Cerretti, D.P., Larsen, A., Park, L., March, C., Dower, S., Gillis, S., and Urdal, D. (1984) *Nature* 312, 768-771.

5. Davis, M.M., Cohen, D.I., Nielsen, E.A., Steinmetz, M., Paul, W.E., and Hood, L. (1984) *Proc. Natl. Acad. Sci. USA* 81, 2194-2198.

6. Dukovich, M., Wano, Y., Thuy, L.t.B., Katz, P., Cullen, B.R., Kehrl, J.H., and Greene, W.C. (1987) *Nature* 327, 518-521.

7. Fujii, M., Sugamura, K., Sano, K., Nakai, M., Sugita, K., and Hinuma, Y. (1986) *J. Exp. Med.* 163, 550-562.

8. Greene, W.C., Robb, R.J., Svetlik, P.B., Rusk, C.M., Depper, J.M.,and Leonard, W.J. (1985) *J. Exp. Med.* 162, 363-368.

9. Hatakeyama, M., Minamoto, S., Uchiyama, T., Hardy, R.R., Yamada, G., and Taniguchi, T. (1985) *Nature* 318, 467-470.

10. Hedrick, S.M., Cohen, D.I., Nielsen, E.A., and Davis, M.M. (1984) *Nature* 308, 149-153.

11. Hirt, B. (1967) *J. Mol. Biol.* 26, 365-369.

12. Ishida, N., Kanamori, H., Noma, T., Nikaido, T., Sabe, H., Suzuki, N., Shimizu, A., and Honjo, T. (1985) *Nucl. Acids Res.* 13, 7579-7589.

13. Kondo, S., Shimizu, A., Maeda, M., Tagaya, Y., Yodoi, J., and Honjo, T. (1986) *Nature* 320, 75-77.

14. LaPolla, R.J., Mayne, K.M., and Davidson, N. (1984) *Proc. Natl. Acad, Sci. USA* 81, 7970-7974.

15. Leonard, W.J., Depper, J.M., Crabtree, G.R., Rudikoff, S., Pumphrey, J., Robb, R.J., Kronke, M., Svetlik, P.B., Peffer, N.J., Waldmann, T.A., and Greene, W.C. (1984) *Nature* 311, 626-631.

16. Leonard, W.J., Depper, J.M., Kanehisa, M., Kronke, M., Peffer, N.J., Svetlik, P.B., Sullivan, M., and Greene, W.C. (1985) *Science* 230, 633-639.

17. Lowenthal, J.W., Zubler, R.H., Nabholz, M., and MacDonald, H.R. (1985) *Nature* 315, 669-672.

18. Lubbert, H., Hoffman, B.J., Snutch, T.P., van Dyke, T., Levine, A.J., Hartig, P.R., Lester, H.A., and Davidson, N. (1987) *Proc. Natl. Acad. Sci. USA* 84, 4332-4336.

19. Nikaido, T., Shimizu, N., Ishida, N., Sabe, H., Teshigawara, K., Maeda, M., Uchiyama, T., Yodoi, J., and Honjo, T. (1984) *Nature* 311, 631-635.

20. Oppenheim, J.J., Kovacs, E.J., Matsushima, K., and Durum, S.K. (1986) *Immunology Today* 7, 45-56.

21. Robb, R.J., Munck, A., and Smith, K.A. (1981) *J. Exp. Med.* 154, 1455-1474.

22. Robb, R.J., Greene, W.C., and Rusk, C.M. (1984) *J. Exp. Med.* 160, 1126-1146.

23. Robb, R.J., Rusk, C.M., Yodoi, J., and Greene, W.C. (1987) *Proc, Natl. Acad. Sci. USA* 84, 2002-2006.

24. Sharon, M., Klausner, R.D., Cullen, B.R., Chizzonite, R., and Leonard, W.J. (1986) *Science* 234, 859-863.

25. Smith, K.A. (1980) *Immunol. Rev.* 51, 337-357.

26. Smith, K.A. (1988) *Immunology Today* 9, 36-37.

27. Teshigawara, K., Wang, H.-M., Kato, K., and Smith, K.A. (1987) *J. Exp. Med.* 165, 223-238.

28. Tsudo, M., Kozak, R.W., Goldman, C.K., and Waldmann, T.A. (1986) *Proc. Natl. Acad. Sci. USA* 83, 9694-9698.

29. Tsudo, M., Kozak, R.W., Goldman, C.K., and Waldmann, T.A. (1987a) *Proc. Natl. Acad. Sci. USA* 84, 4215-4218.

30. Tsudo, M., Goldman, C.K., Bongiovanni, K.F., Chan, W.C., Winton, E. F., Yagita, M., Grimm, E.A., and Waldmann, T.A. (1987b) *Proc. Natl. Acad. Sci. USA* 84, 5394-5398.

31. Waldmann, T.A. (1986) *Science* 232, 727-732.

32. Wang, A., Lu, S.D., and Mark, D. (1984) *Science* 224, 1431-1433.

33. Wang, H.-M., and Smith, K.A. (1987) *J. Exp. Med.* 166, 1055-1069.

34. Weissman, A.M., Harford, J.B., Svetlik, P.B., Leonard, W.L., Depper, J.M., Waldmann, T.A., Greene, W.C., and Klausner, R.D. (1986) *Proc. Natl. Acad. Sci. USA* 83, 1463-1466.

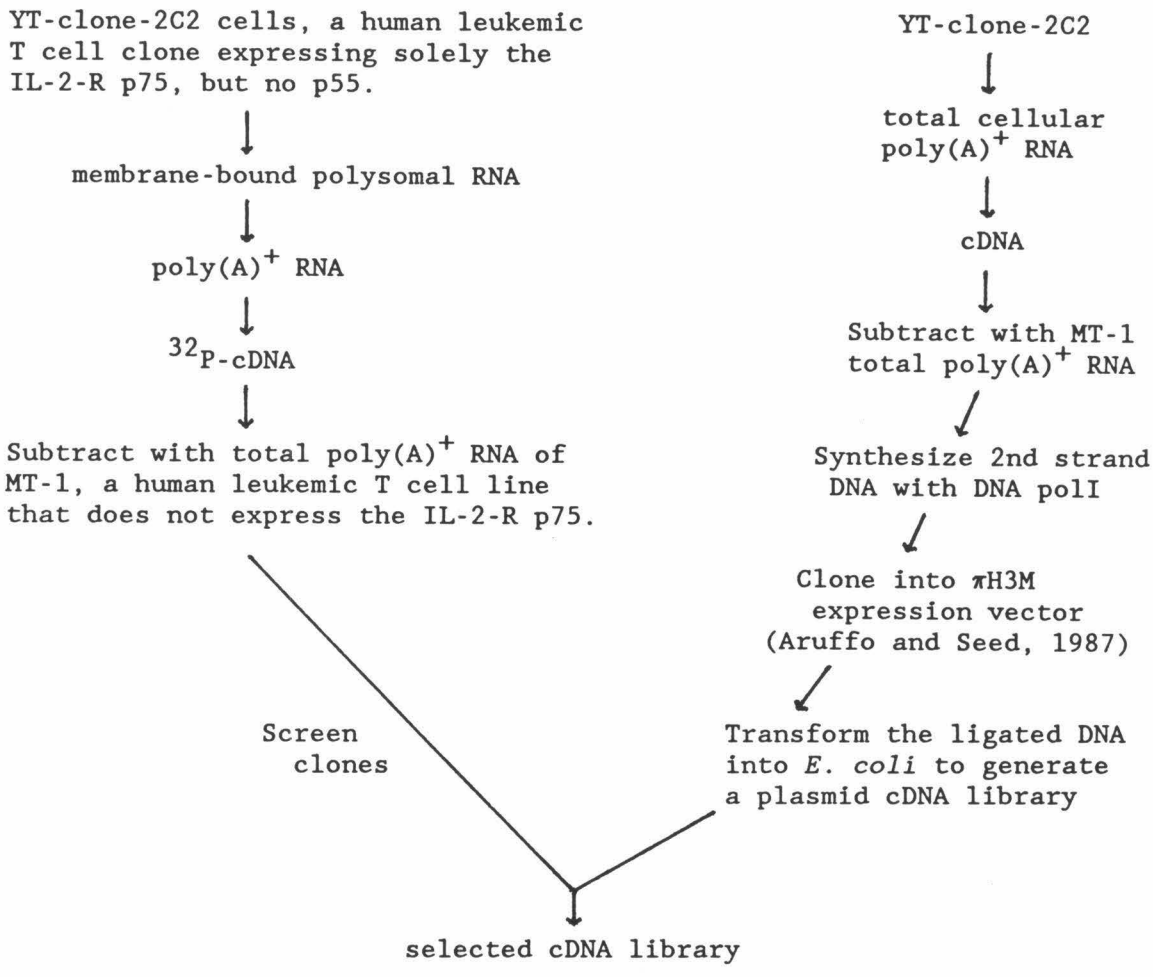


Figure 1. Strategy of clone isolation.

PROPOSITION 2

Cloning of cDNA(s) encoding the negative regulatory protein(s) that modulate expression of human c-*myc* by a new genetic approach.

ABSTRACT

Regulation of transcription from the proto-oncogene c-*myc* appears to play an important role in normal cellular proliferation and the genesis of diverse tumors. A negative regulator has been identified within a DNA segment upstream of the human c-*myc* gene that can inhibit expression from its promoters. Moreover, a 15- to 20-bp sequence within this DNA segment binds a nuclear factor (or factors) from extracts of human Hela cell or mouse ID13 cell nuclei (Hay et al., 1987). It is therefore interesting to investigate this putative sequence-specific DNA-binding protein (or proteins) that binds to the negative element and represses the c-*myc* gene. I propose a new genetic approach to directly isolate cDNA clones encoding the negative regulatory DNA-binding protein(s) of human c-*myc* without prior purication of the protein(s). If successful, this system would represent a powerful general approach for the isolation of genes for eukaryotic sequence-specific DNA-binding proteins, such as tissue-specific transcription factors.

INTRODUCTION

Alterations in the structure and/or expression of the proto-oncogene c-*myc* as a consequence of chromosome translocation are well studied in most murine plasmacytomas and human Burkitt lymphomas (for review, see Cory, 1986; Cole, 1986). The mechanisms that regulate the c-*myc* gene have become central to understanding the role of c-*myc* in cell transformation and in the response of cells to mitogens. Unlike the *ras* oncogenes, which are activated as the result of mutations within the protein-coding regions, no amino acid changes are required for c-*myc* activation. Many investigations, therefore, have pointed to quantitative differences in the levels of c-*myc* expression in tumor vs. normal cells. Expression of c-*myc* is increased during the response of cells to growth factors (Greenberg and Ziff, 1984; Greenberg et al., 1985; Dean et al., 1986; Levine et al., 1986) and decreased during differentiation of cells to a nonproliferative state (Grosso and Pitot, 1985; Bentley and Groudine, 1986; Endo and Nadal-Ginard, 1986; Ramsay et al., 1986; Yarden and Kimchi, 1986), suggesting that c-*myc* protein is an important component in the regulatory networks associated with normal cell proliferation. Since the c-*myc* protein has been shown to be a DNA-binding protein and associated with the cell nuclear matrix (Donner et al., 1982; Eisenman et al., 1985; Watt et al., 1985), it is also possible that c-*myc* is involved in the transcriptional control of other cellular genes which in turn regulate cell proliferation.

The available data are consistent with the view that chromosomal translocations and proviral insertion lead to deregulation or constitutive expression of the c-*myc* gene. Several models have been proposed to

explain how these DNA rearrangements disrupt normal c-*myc* regulation. First, the c-*myc* gene has been shown to be regulated at the transcriptional level since serum or growth factors can stimulate quiescent cells to increase the c-*myc* transcription rate significantly (Greenberg and Ziff, 1984; Greenberg et al., 1985; Dean et al., 1986; Levine et al., 1986), i.e. the positive transcriptional control model. Second, the c-*myc* gene has been proposed to be under negative transcriptional control via a putative repressor which binds to the 5'-flanking region or within the large first untranslated exon (Dunnick et al., 1983; Leder et al., 1983), i.e. the negative transcriptional control model. Third, it has been suggested that modulation of RNA stability plays an important role in the regulation of c-*myc* RNA level (Blanchard et al., 1985; Dani et al., 1985; Dony et al., 1985; Knight et al., 1985; Piechaczyk et al., 1985), i.e. the post-transcriptional modulation model. Fourth, c-*myc* has been postulated to regulate its own level of expression through a feedback mechanism, which would account for the suppression of the normal allele in tumor cells with activated c-*myc* genes (Dunnick et al., 1983; Leder et al., 1983; Nishikura et al., 1983), i.e. the feedback regulation model. Fifth, Saito et al. (1983) have found that a region of the nucleotide sequences in the first exon was significantly complementary to a region of the second exon. The mRNA from the untranslocated c-*myc* gene could form a stable stem-loop structure where the AUG initiation codon would be located within the loop. They proposed that such a secondary structure could render the AUG initiator inaccessible for translation. Thus, displacement of the first exon from the translocated c-*myc* gene would produce higher levels

of protein from a given amount of RNA, i.e. the translational control model. Yet, there is no consensus model that can account for all of these findings and interpretations.

It is clear that the expression of the c-*myc* gene is regulated in part at the transcriptional level. Recently, a negative transcriptional control element has been identified in the murine c-*myc* gene 5'-flanking region, located 428-1188 bp 5' of the first c-*myc* promoter (Remmers et al., 1986). This element is likely to inhibit gene expression at the transcriptional level but not at the post-transcriptional level since it is located outside the transcription unit of the c-*myc* gene. Interestingly, this element can act, in either orientation, upon a heterologous promoter (the SV40 promoter in the pSV2CAT plasmid) at a distance of ≈1.7 kb 3' of this promoter to inhibit gene expression (Remmers et al., 1986). Therefore, it has been proposed that there is a negative enhancer, termed as "dehancer", located upstream of the murine c-*myc* gene. Strikingly, a negative regulator for the human c-*myc* gene has also been identified very recently (Hay et al., 1987). This negative regulatory element of human c-*myc* resides within a segment (-293 to -353) upstream of the gene that can inhibit expression from its promoters (P1 and P2), as shown in Figure 1A (page 107). Furthermore, they have demonstrated that a 15- to 20-bp sequence within this DNA segment (the negative element) binds a nuclear factor (or factors) from extracts of human Hela cell nuclei. In addition, they found the same binding site was occupied when this DNA segment was exposed to nuclear extracts from a mouse cell line ID13, suggesting that a common nuclear factor (or factors) is involved in the inhibition of transcription from

the c-*myc* gene.

A number of eukaryotic genes whose expressions correspond to extracellular inducing signals are under the control of juxtaposed transcriptional enhancers and the negative regulators. In each case, the negative regulators are effective only in cellular conditions where the putative negative *trans*-acting factors (repressor-like molecules) are active (Chambon et al., 1984; Gonzalez and Nebert, 1985; Goodbourn et al., 1986; Nir et al., 1986). One good example is the negative regulator that represses transcription from the human β-interferon gene in the absence of an inducing agent (Goodbourn et al., 1986). This negative regulatory mechanism has been further supported by genomic DNAase I footprinting experiments (Zinn and Maniatis, 1986). Similarly, there may be settings in which the positive regulators of the c-*myc* genes are controlled by the adjacent negative regulators. There are several similarities between these two negative regulators (β-interferon and c-*myc* genes) described above. First, both negative regulatory elements reside downstream of enhancer elements and upstream of their promoters. Second, activation of these promoters may involve both the removal of a repressor (or repressors) and the stimulation of a positive regulatory factor (or factors) so that enhancer elements can override the negative elements in the induced state. Third, both negative regulatory elements can repress transcription from a heterologous promoter (for instance, the SV40 promoter). Fourth, evidently, the nucleotide sequences that bind nuclear factors within both negative elements are conserved (Figure 1B, page 107). In light of these findings, it seems likely that at least some of the negative transcription-

al regulators in mammalian cells may share a common mechanism that a sequence-specific DNA-binding protein(s) directly binds to the negative element of the repressable genes and inhibits transcription.

SPECIFIC AIMS

The primary objectives of this proposal are to isolate cDNA(s) encoding the putative negative regulatory protein(s) responsible for the inhibition of transcription from the human c-*myc* gene, investigate the structure of the protein deduced from the cDNA(s), and further study the regulation of this gene (repressor) expression in normal mammalian cells. This should lead, in the long term, to an understanding of the molecular basis of regulation of transcription from the proto-oncogene c-*myc* and thus contribute to the elucidation of the mechanisms governing the control of normal cellular proliferation and in the genesis of diverse tumors.

METHODS AND PROCEDURES TO BE USED IN PROPOSED PROJECT

I propose two major steps to directly clone cDNA clones encoding the c-*myc*-specific negative regulatory protein(s) by a new genetic approach without prior purification of the protein(s). First, I will construct a cDNA library from human Hela cells by using a redesigned λgt11 system (see below). Second, I will develop a genetic circuit in *E. coli* to serve as an *in vivo* DNA-binding assay (see below). The expression vector λgt11 will be modified by replacing the entire *lacZ* gene with the *trpE* expression cassette of the expression vector pSHNB6, a derivative of the expression vector pATH1, as described (Tanese et

al., 1985). There are several reasons for redesigning this expression vector. In the λgt11 system, the foreign amino acid sequences are fused to the carboxyl terminus of the very large β-galactosidase protein. There is some danger that any segments of the exogenous protein that can recognize DNA will be buried. I therefore propose to use the *trpE* expression cassette (Tanese et al., 1985). In this case, the fusion protein will contain only the first 18 amino acids of *trpE*, then the foreign protein. Therefore, the probability of functional domains in the foreign protein sequence being available for reaction with a DNA sequence will be enhanced. It has been demonstrated that the *trpE* expression cassette of the expression vector pSHNB6 can express high levels of enzymatically active reverse transcriptase that closely resemble the authentic enzyme and are stable in *E. coli* (Roth et al., 1985; Tanese et al., 1985). Furthermore, the *trpE* expression cassette can be induced in the absence of tryptophan by the addition of indole-acrylic acid so that this system can be applied to enrich sequence-specific DNA-binding proteins in *E. coli*.

The proposed *in vivo* DNA-binding assay is designed as follows. I will install the human c-*myc* negative regulatory element (termed NRE) sequence (as shown in Figure 1B, page 107) into the promoter region, between the Pribnow box and the transcription start site, of the *lacZ* gene in the exact position normally occupied by the *lac* operator. The c-*myc* NRE DNA fragment will be prepared by oligonucleotide synthesis on both strands. I anticipate that the expression of β-galatosidase will then be under the control of the putative NRE-binding protein(s) (see below). Lac^+ bacterial colonies or phage plaques are blue in the

presence of X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside). The rationale of this expression system is that, in a bacterial host containing the plasmid with the modified *lacZ* gene, if the λ/*trpE* vector containing the cDNA encodes and expresses a functional NRE-binding protein, then this protein will recognize the NRE sequence in the promoter region of the *lacZ* gene and inhibit the transcription of this gene just as the *lacZ* gene is repressed by the operator-bound repressor in the *lac* operon. This λ/*trpE* plaque will then be colorless (clear plaque) in the presence of X-gal whereas other λ/*trpE* plaques without NRE-binding proteins will be blue. Therefore, I could readily screen λ/*trpE* plaques by the color indication and pick up colorless plaques. Whether or not a large scale lysate from this phage plaque contains NRE-binding activity will be tested by biochemical analyses such as gel-mobility-shift assays or DNAase I footprints. It is important to reiterate that the strategy described above using the λ/*trpE* expression cDNA library is to express the putative NRE-binding proteins at a high level so that they will be able to block most or all of the NRE sequence in the *lacZ* expression vectors. Therefore, the background level of β-galatosidase due to insufficient blocking of NRE-binding proteins can be minimized.

Overall, this system consists of one phage λ expression vector containing a cDNA library and one plasmid expression vector containing the *lacZ* indicator gene; therefore, they should be compatible in an *E. coli* host. Moreover, this system retains the advantage of the high efficiency *in vitro* packaging of the λgt11 system for the generation of a representative cDNA library. If successful, this system would

represent a powerful general approach for the isolation of genes for eukaryotic sequence-specific DNA-binding proteins, such as tissue-specific transcription factors.

It should be noted that there are several potential difficulties in this approach. First, it will be hard to clone full length cDNA in the λgt11 like vector if the mRNA encoding the DNA-binding protein is greater than approximately 3 kb. It has been shown very recently, however, that transcription factor SP1 synthesized in *E. coli* from a partial cDNA clone is still functional for sequence-specific DNA binding *in vivo* and *in vitro* (Kadonaga et al., 1987), suggesting that in all cases it may not be strictly necessary to have a full length cDNA for a functional assay to work. Second, it will be difficult to retain DNA-binding activity with a polypeptide from a single cDNA if the functional protein comprises multiple different subunits. However, most (if not all) of the prokaryotic sequence-specific DNA-binding proteins characterized so far are encoded by single genes, such as the λ and *lac* repressors (Miller and Reznikoff, 1980), *araC* (Miller and Reznikoff, 1980) and *lexA* (Brent and Ptashne, 1984) proteins. This is also true for most of the known eukaryotic sequence-specific DNA-binding or regulatory proteins, including the *Gal4* transcriptional regulatory protein (Johnston and Hopper, 1982), the mating type regulatory proteins α1 and α2 (Johnson and Herskowitz, 1985) in yeast; the SV40 large T antigen (Hansen et al., 1981), glucocorticoid receptors in mammals (Miesfeld et al., 1986); and the transcription factors SP1 (Briggs et al., 1986; Kadonaga and Tjian, 1986), CTF (Jones et al., 1987), AP-1 (Lee et al., 1987), nuclear factor 1 (NF1)

(Rosenfeld and Kelly, 1986), and the adenovirus MLTF (Chodosh et al., 1986) from human Hela cells; heat shock HSTFs from *Drosophila* and yeast (Wiederrecht et al., 1987). These results suggest a single cDNA will probably encode a functional NRE-binding protein. Third, a bacterial synthesized polypeptide may not work if the functional DNA-binding protein requires eukaryotic-specific post-translational modifications. However, in a *Xenopus* oocyte microinjection assay, the human adenovirus E1A protein synthesized in *E. coli* is functional (Ferguson et al., 1984). Also, in a protoplast fusion experiment, the human T-cell leukemia virus type I (HTLV-1) $P40^{x}$ protein, a *trans*-activator (tat-1) protein, synthesized in *E. coli* is functional (Giam et al., 1986). I therefore assume that post-translational modifications may not be a problem for this system.

REFERENCES

1. Bentley, D.L., and Grouidine, M. (1986) *Nature* 321, 702-706.

2. Blanchard, J.-M., Piechaczyk, M., Dani, C., Chambard, J.-C., Franchi, A., Pouyssegur, J., and Jeanteur, P. (1985) *Nature* 317, 443-445.

3. Brent, R. and Ptashne, M. (1984) *Nature* 312, 612-615.

4. Briggs, M.R., Kadonaga, J.T., Bell, S.P., and Tjian, R. (1986) *Science* 234, 47-52.

5. Chambon, P., Gaub, M.P., Lepennec, J.P., Dierich, A., and Astinotti, D. (1984) In Labrie, F. and Prouix, L. (eds.) *Endocrinology*. Excerpta Medica, Amsterdam.

6. Chodosh, L.A., Carthew, R.W., and Sharp, P.A. (1986) *Mol. Cell. Biol.* 6, 4723-4733.

7. Cole, M.D. (1986) *Annu. Rev. Genet.* 20, 361-384.

8. Cory, S. (1986) *Adv. Cancer Res.* 47, 189-234.

9. Dani, C., Mechti, N., Piechaczyk, M., Lebleu, B., Jeanteur, P., and Blanchard, J.-M. (1985) *Proc. Natl. Acad. Sci. USA* 82, 4896-4899.

10. Dean, M., Levine, R.A., Ran, W., Kindy, M.S., Sonenshein, G.E., and Campisi, J. (1986) *J. Biol. Chem.* 261, 9161-9166.

11. Donner, P., Greiser-Wilke, I., and Moelling, K. (1982) *Nature* 296, 262-266.

12. Dony, C., Kessel, M., and Gruss, P. (1985) *Nature* 317, 636-639.

13. Dunnick, W., Shell, B.E., and Dery, C. (1983) *Proc. Natl. Acad. Sci. USA* 80, 7269-7273.

14. Eisenman, R.N., Tachibana, C.Y., Abrams, H.D., and Hann, S.R. (1985) *Mol. Cell. Biol.* 5, 114-126.

15. Endo, T., and Nadal-Genard, B. (1986) *Mol. Cell. Biol.* 6, 1412-1421.

16. Ferguson, B., Jones, N., Richter, J., and Rosenberg, M. (1984) *Science* 224, 1343-1346.

17. Giam, C.-Z., Nerenberg, M., Khoury, G., and Jay, G. (1986) *Proc. Natl. Acad. Sci. USA* 83, 7192-7196.

18. Gonzalez, F.J., and Nebert, D.W. (1985) *Nucleic Acids Res.* 13, 7269-7288.

19. Goodbourn, S., Hurstein, H., and Maniatis, T. (1986) *Cell* 45, 601-610.

20. Greenberg, M.E., and Ziff, E.B. (1984) *Nature* 311, 433-438.

21. Greenberg, M.E., Greene, L.A., and Ziff, E.B. (1985) *J. Biol. Chem.* 260, 14101-14111.

22. Grosso, L.E., and Pitot, H.C. (1985) *Cancer Res.* 45, 847-850.

23. Hansen, U., Tenen, D.G., Livingston, D.M., and Sharp, P.A. (1981) *Cell* 27, 603-612.

24. Hay, N., Bishop, J.M., and Levens, D. (1987) *Genes and Devel.* 1, 659-671.

25. Huynh, T.V., Young, R.A., and Davis, R.W. (1985) In Glover, D. (eds.) *DNA Cloning (Volume I), A Practical Approach.* IRL Press, Oxford, pp. 49-78.

26. Johnson, A.D., and Herskowitz, I. (1985) *Cell* 42, 237-247.

27. Johnston, S. and Hopper, J. (1982) *Proc. Natl. Acad. Sci. USA* 79, 6971-6975.

28. Jones, K.A., Kadonaga, J.T., Rosenfeld, P.J., Kelly, T.J., and Tjian, R. (1987) *Cell* 48, 79-89.

29. Kadonaga, J.T., and Tjian, R. (1986) *Proc. Natl. Acad. Sci. USA* 83, 5889-5893.

30. Kadonaga, J.T., Carner, K.R., Masiarz, F.R., and Tjian, R. (1987) *Cell* 51, 1079-1090.

31. Knight, E.Jr., Anton, E.D., Fahey, D., Friedland, B.K., and Jonak, G.J. (1985) *Proc. Natl. Acad. Sci. USA* 82, 1151-1154.

32. Leder, P., Battey, J., Lenoir, G., Moulding, C., Murphy, W., Potter, H., Stewart, T., and Taub, R. (1983) *Science* 222, 765-771.

33. Lee, W., Mitchell, P., and Tjian, R. (1987) *Cell* 49, 741-752.

34. Levine, R.A., McCormack, J.E., Buckler, A., and Sonenshein, G.E. (1986) *Mol. Cell. Biol.* 6, 4112-4116.

35. Miesfeld, R., Rusconi, S., Godowski, P.J., Maler, B.A., Okret, S., Wikstrom, A.-C., Gustafsson, J.-A., and Yamamoto, K.R. (1986) *Cell* 46, 389-399.

36. Miller, J.H. and Reznikoff, W.S. (1980) *The Operon*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

37. Nir, U., Walker, M.D., and Rutter, W.J. (1986) *Proc. Natl. Acad. Sci. USA* 83, 3180-3184.

38. Nishikura, K., Abbas, R., Erikson, J., Watt, R., Rovena, G., Croce, C.M. (1983) *Proc. Natl. Acad. Sci. USA* 80, 4822-4826.

39. Piechaczyk, M., Yang, Y.-Q., Blanchard, J.-M., Jeanteur, P., and Marcu, K.B. (1985) *Cell* 42, 589-597.

40. Ramsay, R.G., Ikeda, K., Rifkind, R.A., and Marks, P.A. (1986) *Proc. Natl. Acad. Sci. USA* 83, 6849-6853.

41. Remmers, E.F., Yang, J.Q., and Marcu, K.B. (1986) *EMBO J.* 5, 899-904.

42. Rosenfeld, P., and Kelly, T. (1986) *J. Biol. Chem.* 261, 1398-1408.

43. Roth, M.J., Tanese, N., and Goff, S.P. (1985) *J. Biol. Chem.* 260, 9326-9335.

44. Saito, H., Hayday, A.C., Wiman, K., Hayward, W.S., and Tonegawa, S. (1983) *Proc. Natl. Acad. Sci. USA* 80, 7478-7480.

45. Tanese, N., Roth, M., and Goff, S.P. (1985) *Proc. Natl. Acad. Sci. USA* 82, 4944-4948.

46. Watt, R.A., Shatzman, A.R., and Rosenberg, M. (1985) *Mol. Cell. Biol.* 5, 448-456.

47. Wiederrecht, G., Shuey, D.J., Kibbe, W.A., and Parker, C.S. (1987) *Cell* 48, 507-515.

48. Yarden, A., and Kimchi, A. (1986) *Science* 234, 1419-1421.

49. Zinn, K., and Maniatis, T. (1986) *Cell* 45, 611-618.

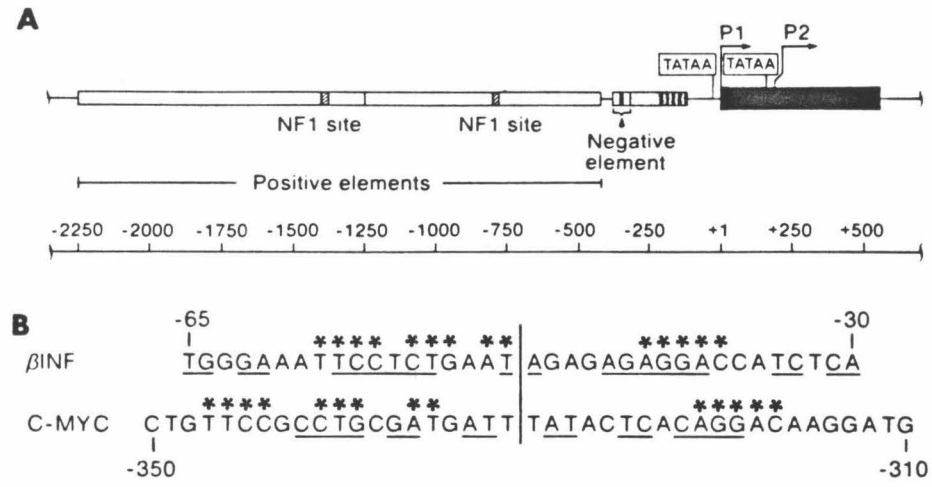


Figure 1. Elements that control transcription from the human c-*myc* gene. (A) Location of the negative and positive regulatory elements. Horizontal arrowheads indicate the sites for initiation of transcription controlled by the promoters P1 and P2. The first untranslated exon is shown as a solid box. Two regions where nuclear factor 1 (NF1) might bind are marked by diagonal hatching boxes. (B) Similarities between the nucleotide sequences in the negative regulatory elements for the human c-*myc* and β-interferon genes. Nucleotides contributing to dyad symmetry are underlined. Asterisks (*) denote identities between the two sequences. (The details of this figure have been taken from Hay et al., 1987, and the pertinent literature.)

PROPOSITION 3

Interferon may induce a negative regulatory protein (or proteins) that is responsible for transcriptional repression of interferon-inducible genes in human cells.

ABSTRACT

Interferons (IFNs) are cell-secreted protein factors that are not only involved in the defense against virus infections, but play a key role in the regulation of cell proliferation and differentiation. Many cellular genes have now been identified as being specifically activated by IFNs and, therefore, they are possible mediators of the complex biological function of IFNs. In human fibroblasts, the induction of some of these genes by IFN (type I) is followed by a decline in the transcription rate of the genes and accompanied by an inability of further IFN exposure to reinduce transcription (Larner et al., 1986). This negative regulatory event is probably due to the synthesis of an IFN-induced regulatory protein (or proteins) that is responsible for transcriptional repression of these genes. I propose a biochemical approach to isolate the putative IFN-induced negative regulatory protein(s) and clone the gene(s) coding for them. If successful, this should contribute to the elucidation of the mechanisms governing the control of IFN-inducible gene regulation.

INTRODUCTION

One of the central issues in modern biology is the control of gene expression. An important question in this regard is how various control molecules, such as specific polypeptides, act on cell surface receptors to control the expression of specific genes which in turn regulate cell growth and function. The interferons (IFNs) are a group of polypeptides that exert potent growth inhibitory and antiviral activities on cells with which they interact (for a review, see Lengyel, 1982). The binding of IFNs to specific high-affinity cell surface receptors results in an induction of particular mRNAs and of the corresponding proteins (Revel and Chebath, 1986). The transcriptional response to interferon is rapid; at least seven genes (namely, 1-8; IND-1; *ISG*-15, -54 and 56; MT-1 and -2) achieve maximum transcription rates within two hours of IFN exposure (Friedman et al., 1984; Larner et al., 1986; Levy et al., 1986; Morris and Huang, 1987; Reich et al., 1987). In human fibroblasts, the induction of IND-1 and *ISG*-54 genes by interferon is followed by a programmed decline in the transcription rate of IND-1 and *ISG*-54 genes and accompanied by an inability of further IFN exposure to reinduce transcription (Larner et al., 1986). A similar transient transcriptional induction has also been found for the *ISG*-15, -54, and -56 genes (Levy et al., 1986; Reich et al., 1987). This negative regulatory event, termed desensitization, requires the continuous synthesis of highly labile proteins because it can be reversed by cycloheximide (Larner et al., 1986; Levy et al., 1986). Although desensitization is known to occur after treatment of many cell types with a variety of hormones and neurotransmitters (Mukherjee et al., 1975; Terasaki et

al., 1978), this is the first instance of desensitization shown to act at the tran-scriptional level for defined mRNA products. Larner et al. (1986) have recently shown that desensitization of transcription of the two IND genes requires newly synthesized protein(s) and is not based solely on a loss of surface IFN receptors. Furthermore, it has been previously claimed that there are no large scale changes in cytoplasmic alkalini-zation, Ca^{2+} flux, and phosphoinositide turn-over during early (0-30 min) treatment of human cells with IFN (Mills et al., 1985). These mechanisms have been variously implicated in the generation of second messengers mediating the mitogenic activation of cells by growth factors (Macara, 1985; Majerus et al., 1984). Therefore, Larner et al. (1986) have proposed the straightforward explanation that an IFN-induced regulatory protein is responsible for negative autoregulation.

What might be the mechanism of transcriptional repression by the negative regulatory protein(s)? The negative regulatory protein(s) may be either an enzyme (for instance, a kinase) or a protein directly involved in transcription. Two IFN-induced enzymes have been characterized (Baglioni, 1979; Lengyel, 1982; McMahon and Kerr, 1983). First, a double-stranded RNA-dependent protein kinase that phosphorylates the α-subunit of protein synthesis initiation factor 2 (eIF2) and blocks translation. This kinase does not appear to be the right candidate since desensitization requires continuous protein synthesis. A second enzyme, 2',5'-oligoadenylate (2-5A) synthetase, is activated by double-stranded RNA, and its product activates a latent endoribonuclease, the 2-5A-dependent RNAase (Wreschner et al., 1981). It has been shown that inhibiting the RNAase by the inhibitor $CH_3Sp(A2'p)_2A2'pp3'OCH_3$ (Watling

et al., 1985), a 2-5A analogue capable of inhibiting the activation of the 2-5A-dependent RNAase, has no major effect on the turnover of rapidly induced mRNAs in the mitogenically stimulated mouse 3T3 and L cells (Porter et al., 1986). This implies that the putative negative regulatory protein(s) is probably not a sequence-specific RNAase. While it is still possible that some other enzymatic activity causes negative regulation, we think it more likely that negative regulation is due to a sequence-specific DNA-binding protein(s) that directly binds to *cis*-acting sequences of the repressible genes, just as the *lac* repressor binds to its cognate operator (Miller and Reznikoff, 1980), and blocks transcription. Similarly, it has been shown that the tissue specificity of the immunoglobulin heavy-chain (IgH) enhancer is negatively controlled (Imler et al., 1987). Blocking protein synthesis with cycloheximide in fibroblasts, in which the IgH gene is normally inactive, stimulates transcription of an integrated IgH gene, suggesting that these cells contain highly labile repressor-like proteins (Ishihara et al., 1984). Furthermore, a number of eukaryotic genes have been shown to be negatively regulated, including those from the cellular genes rat insulin 1 (Laimins et al., 1986; Nir et al., 1986), human β-interferon (Goodbourn et al., 1986), murine and human *c-myc* (Remmers et al., 1986; Hay et al., 1987), and from the SV40 (Borrelli et al., 1984; Velcich and Ziff, 1985), polyomavirus (Borrelli et al., 1984; Cremisi and Babinet, 1986), and murine sarcoma virus (Gorman et al., 1985). In light of these results, it seems reasonable that negative regulation of eukaryotic gene expression may be a common mechanism. It has been proposed that these negative regulatory elements function by interfering

with the interaction of enhancers with ubiquitous stimulatory *trans*-acting factors. Moreover, once the *cis*-acting sequence responsible for the putative IFN-induced negative regulator(s) has been identified (see below), I should be able to verify whether the nature of the negative regulator(s) is a sequence-specific DNA-binding protein(s) by biochemical analyses such as gel-mobility-shift assays, or DNAase I footprinting.

SPECIFIC AIMS

The primary objectives of this proposal are to isolate the putative IFN-induced negative regulatory protein(s), clone the gene(s) coding for them, and further study the regulation of this gene expression in mammalian cells. This should lead, in the long term, to an understanding of the molecular basis of IFN-dependent transcriptional desensitization and thus contribute to the elucidation of the mechanisms governing the control of IFN-inducible gene regulation.

METHODS AND PROCEDURES TO BE USED IN PROPOSED PROJECT

1. Delineation of *cis*-acting sequences responsible for the putative negative regulator(s). Three IFN-inducible human genes have been previously isolated and characterized (Levy et al., 1986; Reich et al., 1987). These genes, termed *ISG*-15, -54, and -56, are inducible by human type I IFN (-α and -β) and subject to desensitization. The 5' portion of the gene *ISG*-54 can drive a heterologous gene lacking an active promoter in the same IFN-dependent manner as that of the gene *ISG*-54, indicating that the regulatory elements for both the positive

and negative arms of IFN-mediated transcription lie within the 5' portion of *ISG*-54 (Levy et al., 1986). Recently, the positive regulatory region of the *ISG*-54 gene has been shown to reside within 122 nucleotides upstream of the transcription initiation site (the cap site) (Levy et al., 1988). Further, an IFN-inducible enhancer element has been recently identified in the 5' flanking region of the *ISG*-15 gene, close to the CCAAT box. This element is highly conserved between the *ISG*-15 and -54 genes (Reich et al., 1987).

It should now be possible, using either deletions or saturation mutagenesis, to identify the *cis*-acting sequence responsible for the negative regulation of these genes. I will use *ISG*-54 as a working system for this purpose. The reported *ISG*-54/E1B fusion vector (Levy et al., 1986) can be used as an expression system. Alternatively, this promoter can be fused to the bacterial chloramphenical acetyltransferase (CAT) gene which is a widely used reporter gene for expression studies (Gorman et al., 1982). Sequential deletion mutants of the 5' flanking DNA of *ISG*-54 will be constructed *in vitro*. In addition, the conserved region between *ISG*-15 and -54 (which has been shown by D. Levy to be the site of a protein footprint after IFN-induction) will be subjected to multiple base-pair changes to alter it and to construct plasmids that differ in positive or negative regulation. Previously, Levy et al. (1986) have described an assay system for determining both the tran-scriptional efficiency and specificity of the recombinant vector *ISG*-54/E1B in human cells. In a similar fashion, I propose to introduce the mutant plasmids into human fibroblast (FS2) cells by a standard DNA-mediated gene transfer technique (for instance, the DEAE-

dextran procedure which I have already used). Subsequently, transfected cells will be treated with IFN-α in the absence or presence of cycloheximide. Cytoplasmic RNA will be extracted and analyzed by the RNAase protection assay as described (Levy et al., 1986; Melton et al., 1984). Alternatively, the IFN-induced desensitization (or repression) of transcription in transfected cells could be analyzed by nuclear runoff assays (Larner et al., 1986) using the E1B DNA of the plasmid to monitor the transcription signal. This approach will probably lead us to identify not only the negative sequence element but also the positive element responsible for transcriptional induction. In order to examine the mechanism by which the negative element functions, the identified *cis*-acting negative element could be fused to the HSV-1 *tk* gene (McKnight and Gavis, 1980) containing its native promoter. Thus, I could examine whether the *cis*-acting negative element functions on a heterologous promoter in an IFN-dependent manner. As a positive control, the 5′ flanking DNA of *ISG*-54 (approximately from -800 to -34 bp) would be fused to the HSV-1 *tk* gene containing its native promoter.

2. Molecular cloning of the gene(s) encoding the negative regulatory protein(s). Once the *cis*-acting interferon-inducible negative element (termed INE) has been identified, it should be possible to isolate the protein(s) that binds to the INE by using the radioactive-labeled INE sequence as a probe (see below). Previous results indicate that the putative INE-binding protein(s) is inducible by type I IFN and causes IFN-dependent desensitization (McMahon and Kerr, 1983). An important question therefore is what role the INE-binding protein(s) plays in initiating the decline in the IFN-induced

transcription of these genes (*ISG*-15, -54, -56). Moreover, since these genes (*ISG*-15, -54, -56) can be induced by type I IFN but not type II IFN (-γ), does the putative INE-binding protein(s) only respond to type I IFN? Finally, how is this INE-binding protein(s) induced by IFNs (type I)? Does this induction also require other IFN-regulated transcription factors, such as the positive transcription regulator(s)?

One approach to answering these questions is the molecular cloning and characterization of the gene(s) coding for INE-binding protein(s). I propose to clone the gene(s) by a biochemical approach. One of the most effective modern biochemical approaches for purifying DNA-binding proteins is sequence-specific DNA affinity chromatography. A purification procedure for the putative INE-binding protein(s) will be designed by analogy with that used for the successful isolation of transcription factors SP1 and CTF (Briggs et al., 1986; Kadonaga and Tjian, 1986; Jones et al., 1987). I propose to prepare nuclear extracts from the IFN (type I) treated human fibroblast (FS2) cells in the presence of protease inhibitors. The several enrichment and purification steps will include ammonium sulfate precipitation, gel filtration (Briggs et al., 1986), ion exchange chromatography (if needed), but especially DNA affinity chromatography (in the presence of excess poly d[I-C]), with tandem copies of the double-stranded INE DNA sequences covalently affixed to Sepharose (Kadonaga and Tjian, 1986). Purification will be monitored by gel-mobility shift assays and/or DNAase I footprints. The highly purified preparation of INE-binding proteins can be analyzed by SDS gel electrophoresis and the molecular weight of the predominant protein determined. The question of whether

the purified INE-binding proteins can induce transcription repression of an IFN-inducible gene will be examined by *in vitro* transcription analysis using the *ISG*-54/E1B fusion vector as a template. It has been determined that constitutive transcription of the *ISG*-54 gene takes place with extracts from uninduced cells (Levy et al., 1986).

Once the putative INE-binding protein(s) has been purified to near homogeneity by sequence-specific DNA affinity chromatography, it should be possible to obtain a partial amino acid sequence and prepare appropriate oligonucleotide probes which will allow us to clone the gene from a human fibroblast cell λgt10 cDNA library (Huynh et al., 1985). Alternatively, purified INE-binding protein(s) will allow us to generate antibodies against this protein, in order to clone the gene from a human fibroblast cell λgt11 cDNA library (Huynh et al., 1985).

Of course, isolation of a protein that binds to a *cis*-acting regulatory sequence permits the inference, but not the rigorous conclusion that the protein functions as a regulator. If the *cis*-acting sequences for positive and negative (INE) regulation overlap, then we may not have a strong inference as to whether we have isolated the positive or negative *trans*-acting regulator. To settle these questions, we will have to set up functional assays based on expression systems for the cloned gene. These may be either an *in vitro* transcription system or based on gene transfer into cells in culture. These are important problems, but they are far enough into the future so that I will not describe in detail the several straightforward approaches to their solutions.

REFERENCES

1. Baglioni, C. (1979) *Cell* 17, 255-264.

2. Borrelli, E., Hen, R., and Chambon, P. (1984) *Nature* 312, 608-612.

3. Briggs, M.R., Kadonaga, J.T., Bell, S.P., and Tjian, R. (1986) *Science* 234, 47-52.

4. Cremisi, C., and Babinet, C. (1986) *J. Virol.* 59, 761-763.

5. Friedman, R.L., Manley, S.P., McMahon, M., Kerr, I.M., and Stork, G.-R. (1984) *Cell* 38, 745-755.

6. Goodbourn, S., Hurstein, H., and Maniatis, T. (1986) *Cell* 45, 601-610.

7. Gorman, C.M., Moffat, L.F., and Howard, B.H. (1982) *Mol. Cell. Biol.* 2, 1044-1051.

8. Gorman, C.M., Rigby, P.W.J., and Lane, D.P. (1985) *Cell* 42, 519-526.

9. Hay, N., Bishop, J.M., and Levens, D. (1987) *Genes and Devel.* 1, 659-671.

10. Huynh, T.V., Young, R.A., and Davis, R.W. (1985) In Glover, D. (eds.) *DNA Cloning (Volume I), A Practical Approach.* IRL Press, Oxford, pp.49-78.

11. Imler, J.-L., Lemaire, C., Wasylyk, C., and Wasylyk, B. (1987) *Mol. Cell. Biol.* 7, 2558-2567.

12. Ishihara, T., Kudo, A., and Watanabe, T. (1984) *J. Exp. Med.* 160, 1937-1942.

13. Jones, K.A., Kadonaga, J.T., Rosenfeld, P.J., Kelly, T.J., and Tjian, R. (1987) *Cell* 48, 79-89.

14. Kadonaga, J.T., and Tjian, R. (1986) *Proc. Natl. Acad. Sci. USA* 83,

5889-5893.

15. Laimins, L., Holmgren-Konig, M., and Khoury, G. (1986) *Proc. Natl. Acad. Sci. USA* 83, 3151-3155.

16. Larner, A.C., Chaudhure, A., and Darnell, J.E.Jr. (1986) *J. Biol. Chem.* 261, 453-459.

17. Lengyel, P. (1982) *Annu. Rev. Biochem.* 51, 251-282.

18. Levy, D., Larner, A., Chaudhuri, A., Babiss, L.E., and Darnell, J.E.Jr. (1986) *Proc. Natl. Acad. Sci. USA* 83, 8929-8933.

19. Levy, D., Pine, R., Kessler, D., Reich, N., and Darnell, J.E.Jr. (1988) *Genes and Devel.* 2, in press.

20. Macara, I.G. (1985) *Am. J. Physiol.* 248, C3-C11.

21. Majerus, P.W., Neufeld, E.J., and Wilson, D.B. (1984) *Cell* 37, 701-703.

22. McKnight, S.L., and Gavis, E.R. (1980) *Nucleic Acids Res.* 8, 5931-5948.

23. McMahon, M., and Kerr, I.M. (1983) In Burke, D.C. and Morris, A.G., (eds.) *Interferons: From Molecular Biology to Clinical Applications*. Cambridge University Press, Cambridge, pp. 89-108.

24. Melton, D.A., Kreig, P.A., Rebagliotti, M.R., Maniatis, T., Zinn, K., and Green, M.R. (1984) *Nucleic Acids Res.* 12, 7035-7056.

25. Miller, J.H. and Reznikoff, W.S. (1980) *The Operon*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

26. Mills, G.B., Hannigan, G., Stewart, D., Mellors, A., Williams, B., and Gelfand, E.W. (1985) In Williams, B.R.G. and Silverman, R.H., (eds.) *The 2-5A System: Clinical and Molecular Aspects of the Interferon Regulated Pathway*. Alan R. Liss, NY, pp. 357-367.

27. Morris, S., and Huang, P.C. (1987) *Mol. Cell. Biol.* 7, 600-605.

28. Mukherjee, C., Caron, M.G., and Lefkowitz, R.J. (1975) *Proc. Natl. Acad. Sci. USA* 72, 1945-1949.

29. Nir, U., Walker, M.D., and Rutter, W.J. (1986) *Proc. Natl. Acad. Sci. USA* 83, 3180-3184.

30. Porter, A.C.G., Chernajovsky, Y., Kelly, J.M., Watling, D., Stark, G.R., and Kerr, I.M. (1986) In Friedman, R.M., Merigan, T., and Sreevalsan, T. (eds.) *Interferons as Cell Growth Inhibitors and Antitumor Factors*. Alan R. Liss, NY, pp. 113-120.

31. Reich, N., Evans, B., Levy, D., Fahey, D., Knight, E.Jr., and Darnell, J.E.Jr. (1987) *Proc. Natl. Acad. Sci. USA*, in press.

32. Remmers, E.F., Yang, J.-Q., and Marcu, K.B. (1986) *EMBO J.* 5, 899-904.

33. Revel, M., and Chebath, J. (1986) *Trends Biochem. Sci.* 11, 166-170.

34. Terasaki, W.L., Brooker, G., de Vellis, J., Inglish, D., Hsu, C.-Y., and Moylan, R.D. (1978) *Adv. Cyclic Nucleotide Res.* 9, 33-52.

35. Velcich, A., and Ziff, E. (1985) *Cell* 40, 705-716.

36. Watling, D., Serafinowska, H.T., Reese, C.B., Kerr, I.M. (1985) *EMBO J.* 4, 431-436.

37. Wreschner, D.H., McCauley, J.W., Skehel, J.J., and Kerr, I.M. (1981) *Nature* 289, 414-417.

PROPOSITION 4

The use of a biotin-streptavidin complex to probe the mechanism of transcriptional enhancer function.

ABSTRACT

Transcriptional enhancers were originally identified as long-range activators of gene transcription in many viral genomes and near or within cellular genes of higher eukaryotes. These cellular enhancers were the first DNA sequences found to confer tissue specificity. The available data are consistent with the view that enhancer function is mediated by sequence-specific *trans*-acting factors (presumably proteins) that bind to enhancer elements and stimulate transcription from nearby promoters, *in vivo* and *in vitro*. However, the mechanism by which enhancer elements interact with specific binding factors and lead to the long-range activation of transcription from promoters is not understood for any enhancer. Although many models have been postulated to account for the mechanism of enhancer action, current experimental data could not exclusively distinguish one model from the others (especially between the looping model and the scanning model). I propose to construct a biotin-streptavidin-biotin bridge between an enhancer and an enhancerless promoter and ask whether the enhancer can then activate the promoter. A positive result will support the looping model.

INTRODUCTION

One of the central problems in eukaryotic molecular biology is to understand the control of gene expression. Essentially, two DNA sequence elements (namely, promoters and enhancers) are required for the regulation of gene transcription in eukaryotes. Promoters are required for accurate and efficient initiation of transcription from genes, whereas enhancers are needed to increase the rate of transcription from promoters. Promoters are located immediately upstream of the start site of transcription (for reviews, see Dynan and Tjian, 1985; McKnight and Tjian, 1986), whereas enhancers can act on *cis*-linked promoters at great distances (a few kilobases away) in an orientation-independent manner and can also function downstream from the transcription unit (for a review, see Serfling et al., 1985). Enhancers have been identified in many viral genomes (for reviews, see Gluzman, 1985; Serfling et al., 1985) and near or within a variety of cellular genes of higher eukaryotes (Banerji et al., 1983; Walker et al., 1983; Gillies et al., 1984; Ott et al., 1984; Picard and Schaffner, 1984; Deschamps et al., 1985; Goodburn et al., 1985; Haslinger and Karin, 1985; Theisen et al., 1986; Kawamoto et al., 1988). These cellular enhancers often potentiate transcription in a tissue-specific, and sometimes inducible, fashion.

The available data are consistent with the view that enhancer function is mediated by sequence-specific *trans*-acting factors (presumably proteins) that bind to enhancer elements and stimulate transcription from nearby promoters, *in vivo* (Scholer and Gruss, 1984; Ephrussi et al., 1985; Mercola et al., 1985; Schlokat et al., 1986) and

in vitro (Sassone-Corsi et al., 1985; Wildeman et al., 1984; Sen and Baltimore, 1986; Wildeman et al., 1986; Sealey and Chalkley, 1987; Speck and Baltimore, 1987; Kawamoto et al., 1988). This stimulation has been implicated in increasing density of RNA polymerase II molecules on actively transcribed genes under enhancer control (Weber and Schaffner, 1985; Treisman and Maniatis, 1985). There are generally multiple binding sites for nuclear factors within an enhancer, and enhancers are often present in two or more copies. Although enhancer sequences have been extensively characterized and their binding factors found, the mechanism by which they interact with protein factors and lead to the long-range activation of transcription from promoters is not understood for any enhancer. Many models have been postulated to account for the mechanism of enhancer function, including the looping model; the scanning model; the chromatin structure model; the site-specific gyrase model; and the nuclear address model (Courey et al., 1986; Ptashne, 1986). In light of numerous experiments, the current view of enhancer function is that enhancers act as entry sites for transcription factors that communicate with other promoter-specific components of the transcriptional machinery, including RNA polymerase II, to initiate transcription. This communication could occur mainly in two distinct ways. First, the enhancer binding factors could scan along the DNA double-helix chain until a promoter and/or its transcriptional machinery is encountered (the scanning model). Second, the intervening DNA double-helix chain could loop out to bring the enhancer and associated protein factors into direct contact with the promoter and/or its transcriptional machinery (the looping model). Yet, current experimental data do not

decisively distinguish between these two models.

SPECIFIC AIMS

The goal of this proposal is to design a simple experimental approach to determine the correct or the most suitable model for the mechanism of enhancer function. This should contribute to an understanding of the mechanism(s) governing the control of gene expression.

METHODS AND PROCEDURES TO BE USED IN PROPOSED PROJECT

I propose to measure the transcription activity of a composite transcription unit in which a DNA fragment containing a transcriptional enhancer element will be linked to another DNA fragment containing solely a promoter and a reporter gene, but no enhancer, by a non-DNA molecule. If this linkage is completely non-DNA like and sufficiently bulky (for instance, a big protein), a scanning mechanism would become improbable to account for any transcriptional activation, whereas a looping mechanism should not be significantly impeded by the structure of the connecting bridge. I therefore propose to introduce a biotin-streptavidin-biotin bridge between the enhancer and promoter sequences as a linkage. A positive result, i.e. observation of an enhancing effect, will be strong evidence for a "through-space" model (the looping model). A negative result will suggest that enhancers operate via a "through-the-chain" model (the scanning model).

Streptavidin is a nonglycosylated 60,000-dalton protein, composed of four identical subunits, isolated from *Streptomyces avidinii* (Chaiet and Wolf, 1964). This protein is extremely similar in function to egg

white avidin and binds four biotin molecules per tetramer with a very high affinity (dissociation constant $K_d=10^{-15}M$) (Green, 1975). Unlike egg white avidin, however, streptavidin exhibits minimal nonspecific binding to biological samples at physiological pH (Haeuptle et al., 1983).

The bacterial chloramphenicol acetyltransferase (CAT) coding sequence that is transcribed under the control of the SV40 early promoter and the SV40 3′ RNA processing signals will be used as a transcription unit which is a widely used reporter gene for an expression study (Gorman et al., 1982a). The proposed construction of the composite enhancer/transcription units are depicted in Figure 1 (page 135). Primarily, the DNA molecule containing the CAT gene with the enhancerless SV40 promoter (SV-CAT) and the DNA molecule containing the enhancer segment will each be labeled with biotin at one end (see below), cross-linked with streptavidin, and the complex transfected into mammalian culture cells. CAT enzyme activity will then be assayed with extracts from the transfected cells. Control experiments will be performed as follows: (1) The two biotinylated DNA molecules (the enhancer molecule and the SV-CAT molecule) without streptavidin will be simultaneously transfected into the same culture cells and assayed for CAT enzyme activity. (2) The biotinylated or non-biotinylated DNA molecule with the SV-CAT gene only (without streptavidin nor enhancer) will be transfected into culture cells and assayed for the basal level of CAT gene expression. (3) The biotinylated DNA molecule with the SV-CAT gene alone (without the enhancer molecule) will be incubated with streptavidin and transfected into culture cells, and thus assayed for the effect

of streptavidin itself on CAT gene expression.

Specifically, the commonly used plasmid $pA_{10}CAT$, which contains the SV40 promoter elements but no enhancer sequences (Gorman et al., 1982b), will be linearized by digestion with SalI and PstI restriction endonucleases so that a DNA fragment (≈2 kb SalI-PstI segment) consisting of the SV40 promoter/CAT gene (SV-CAT) only can be isolated and purified after fractionating by agarose gel electrophoresis. In the same manner, a DNA fragment (≈0.6 kb SalI-SphI segment) containing solely the Moloney murine sarcoma virus (MSV) enhancer will be isolated from the plasmid pSM12 (Laimins et al., 1984) and used as a transcriptional enhancer element. This particular enhancer is chosen because the effect of the MSV enhancer in augmenting expression by insertion into $pA_{10}CAT$ has already been studied. The ends (SalI sites) of these two specific DNA fragments that contain the SV-CAT gene and the MSV enhancer will be repaired individually with *E. coli* DNA polymerase I large fragment (Klenow enzyme) in the presence of nucleotides dCTP, dGTP, biotin-11-dUTP, and 2',3'-dideoxyadenosine 5'-triphosphate (see below for an explanation). ^{32}P-labeled-dCTP will also be included in the end-filling reaction to monitor the percentage of incorporation of biotin-11-dUTP. The biotinylated DNA will be separated from unincorporated biotin-11-dUTP by chromatography on Sephadex G-50 or by a G-50 spun column (Maniatis et al., 1982). (Note that phenol extractions should be avoided during DNA preparation as biotinylated DNA will partition to the interface or into the phenol layer.) The purified biotin-end-labeled DNA fragment containing the SV-CAT gene will be incubated with an excess of streptavidin (the molar ratio will

be around DNA:streptavidin = 1:2), separated from unbound streptavidin by chromatography on Sepharose CL-2B, and then incubated with the purified biotin-end-labeled DNA fragment containing the MSV enhancer element. The final enhancer/streptavidin/SV-CAT complex (≈2.6 kb in length) can be separated and isolated from the other DNA mixtures, including any enhancer/streptavidin/enhancer complex (≈1.2 kb in length), or SV-CAT/streptavidin/SV-CAT complex (≈4 kb in length) and other unbound DNA fragments, after fractionating by agarose gel electrophoresis. Although streptavidin has four biotin-binding sites, the available evidence is that they are grouped into two non-interacting sets of two each (Green et al., 1971). It is therefore probable that the above manipulations will lead to a linear DNA-biotin-streptavidin-biotin-DNA configuration. Subsequently, the purified composite enhancer/streptavidin/SV-CAT molecule will be transfected into mammalian cells in culture (for example, mouse L*tk*$^-$ cells or the rat XC cells) using the DEAE-dextran transfection protocol that is sufficiently effective for transient expression with linear DNA in these culture cells (Hu and Davidson, unpublished data) or the electroporation method (Chu and Berg, 1987). After 48-72 hours incubation, CAT enzyme activities will be assayed with extracts from these transfected cells as described (Gorman et al., 1982a). In the meantime, DNA samples for the various control experiments as described above will be prepared, with or without streptavidin, transfected into culture cells using the same transfection protocol, and CAT enzyme activities will be assayed with extracts from the transfected cells in the same fashion. (All of the materials described above are commercially available.)

It should be noted that the restriction endonucleases SalI (which will generate 5'-protruding ends), PstI and SphI (both of which will generate 3'-protruding ends) are used to linearize the DNA because only the 5'-protruding ends can be repaired with Klenow enzyme in the presence of 4 dNTPs. Therefore, they will be filled in with biotin-11-dUTP and leave the other 3'-protruding ends (PstI and SphI sites) unlabeled. Moreover, the terminus of each repaired SalI site (5'...G▼TCGAC...3' / 3'...CAGCT▲G...5') will be a dideoxy nucleotide (ddATP) which will not be capable of ligating with other DNA blunt ends. It will therefore rule out the possibility of ligation between the enhancer and the SV-CAT gene fragments in the transfected cells if some of the ends of the biotin-labeled DNA are not stably bound with streptavidins (see below). This design also prevents the self-ligation between the two free ends of the composite enhancer/ streptavidin/SV-CAT molecule in the transfected cells since it is unlikely to join the PstI site on the one end, with the SphI site on the other end (Figure 1, page 135). However, one might imagine that it is possible to form a big circular DNA molecule in which two copies of the composite enhancer/streptavidin/SV-CAT molecule might join together via the ligation of PstI to PstI sites and SphI to SphI sites. Should this happen in the transfected cells, the enhancer elements and the transcription units (SV-CAT) are still physically separated by streptavidins so that this peculiar molecule should not distort the proposed experiments. Nevertheless, the physical structure (i.e. circular or linear) of the DNA molecule in the transfected cells will be examined after the DNA transfection (see below).

It is crucial to consider the stability of the streptavidin-linked molecules since variation in the stability of this composite molecule will significantly interfere with the interpretation of transcriptional enhancement. Fortunately, because of the extraordinarily high-affinity interaction between biotin and streptavidin [K_d=10^{-15}M, (Green, 1975)], it seems reasonable to assume that the biotin-streptavidin linkage will be stable in culture cells and in an *in vitro* transcription system (see below). Furthermore, the configuration of the composite enhancer/streptavidin/SV-CAT molecule in transfected cells will be confirmed by extracting the DNA molecules from transfected cells' nuclei by the Hirt procedure (Hirt, 1967), and analyzing the length and the physical structure of the DNA molecule by electrophoresis on agarose gels.

Alternatively, the transcription activity of this composite enhancer/streptavidin/SV-CAT DNA molecule and the various control DNA constructions can be determined in an *in vitro* transcription system in which the composite DNA molecule presumably would be stable since cell-free transcription extracts have become well defined (Reinberg et al., 1987; Reinberg and Roeder, 1987a and b). Although the *in vitro* enhancing effect of a transcriptional enhancer (SV40 enhancer) has been shown to be limited somewhere between 5- to 15-fold (Scholer and Gruss, 1985), the advantage of using such a highly characterized transcription system justifies the attempt to study the mechanism of enhancer function. Yet, there is no proof that the mechanism of transcriptional stimulation of an enhancer *in vitro* is the same as that *in vivo*. It is still debatable whether the full enhancer function can be achieved *in vitro* with cell-free transcription extracts, or whether some higher-

order nuclear structures are also required. Nevertheless, the first approach I propose to use here is an *in vivo* transcription system and the rest of the problems are far enough into the future so that I will not describe in detail the approach of using an *in vitro* transcription system.

Finally, all of the procedures described above are standard molecular and cellular biology techniques. It does not require any particular reactions or difficult techniques to carry out the proposed project. If the transcription activity (CAT enzyme activity) of such a composite enhancer/streptavidin/SV-CAT DNA molecule is significantly higher than those of the control experiments as described above, it will be strong support for the looping model for the mechanism of enhancer function.

REFERENCES

1. Banerji, J., Olson, L., and Schaffner, W. (1983) *Cell* 33, 729-740.

2. Chu, G., Hayakawa, H., and Berg, P. (1987) *Nucleic Acids Res.* 15, 1311-1326.

3. Chaiet, L., and Wolf, F.L. (1964) *Arch. Biochem. Biophys.* 106, 1-5.

4. Deschamps, J., Meijlink, R., and Verma, I.M. (1985) *Science* 230, 1174-1177.

5. Dynan, W.S., and Tjian, R. (1985) *Nature* 316, 774-778.

6. Ephrussi, A., Church, G.M., Tonegawa, S., and Gilbert, W. (1985) *Science* 227, 134-140.

7. Gillies, S.D., Folsom, V., and Tonegawa, S. (1984) *Nature* 310, 594-597.

8. Gluzman, Y. (ed.) (1985) *Eukaryotic transcription: the role of cis- and trans-acting elements in initiation.* Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

9. Goodburn, S., Zinn, K., and Maniatis, T. (1985) *Cell* 41, 509-520.

10. Gorman, C.M., Moffat, L.F., and Howard, B.H. (1982a) *Mol. Cell. Biol.* 2, 1044-1051.

11. Gorman, C.M., Merlino, G.T., Willkingham, M.C., Pastan, J., and Howard, B.H. (1982b) *Proc. Natl. Acad. Sci. USA* 79, 6777-6781.

12. Green, N.M., Konieczny, L., Toms, E.J., and Valentine, R.C. (1971) *Biochem. J.* 125, 781-791.

13. Green, N.M. (1975) *Adv. Protein Chem.* 29, 85-133.

14. Haeuptle, M.T., Aubert, M.L., Djiane, J., and Kraehenbuhl, J.P. (1983) *J. Biol. Chem.* 258, 305-314.

15. Haslinger, A., and Karin, M. (1985) *Proc. Natl. Acad. Sci. USA* 82,

8572-8576.

16. Hirt, B. (1967) *J. Mol. Biol.* 26, 365-369.

17. Kawamoto, T., Makino, K., Niwa, H., Sugiyama, H., Kimura, S., Amemura, M., Nakata, A., and Kakunaga, T. (1988) *Mol. Cell. Biol.* 8, 267-272.

18. Laimins, L.A., Gruss, P., Pozzatti, R., and Khoury, G. (1984) *J. Virol.* 49, 183-189.

19. Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

20. McKnight, S., and Tjian, R. (1986) *Cell* 46, 795-805.

21. Mercola, M., Goverman, J., Mirell, C., and Calame, K. (1985) *Cell* 36, 403-411.

22. Ott, M.O., Sperling L., Herbomel, P., Yaniv, M., and Weiss, M.C. (1984) *EMBO J.* 3, 2505-2510.

23. Picard, D., and Schaffner, W. (1984) *Nature* 307, 80-82.

24. Reinberg, D., Horikoshi, M., and Roeder, R.G. (1987) *J. Biol. Chem.* 262, 3322-3330.

25. Reinberg, D., and Roeder, R.G. (1987a) *J. Biol. Chem.* 262, 3310-3321.

26. Reinberg, D., and Roeder, R.G. (1987b) *J. Biol. Chem.* 262, 3331-3337.

27. Sassone-Corsi, P., Wildeman, A.G., and Chambon, P. (1985) *Nature* 313, 458-463.

28. Schlokat, U., Bohmann, D., Scholer, H., and Gruss, P. (1986) *EMBO J.* 5, 3251-3258.

29. Scholer, H.R., and Gruss, P. (1984) *Cell* 36, 403-411.

30. Scholer, H.R., and Gruss, P. (1985) *EMBO J.* 4, 3005-3013.

31. Sealey, L., and Chalkley, R. (1987) *Mol. Cell. Biol.* 7, 787-798.

32. Sen, R., and Baltimore, D. (1986) *Cell* 46, 705-716.

33. Serfling, E., Jasin, M., and Schaffner, W. (1985) *Trends Genet.* 1, 224-230.

34. Speck, N.A., and Baltimore, D. (1987) *Mol. Cell. Biol.* 7, 1101-1110.

35. Treisen, M., Stief, A., and Sippel, A.E. (1986) *EMBO J.* 5, 719-724.

36. Treisman, R., and Maniatis, T. (1985) *Nature* 315, 72-75.

37. Walker, M.D., Edlund, T., Boulet, A.M., and Rutter, W.J. (1983) *Nature* 306, 557-561.

38. Weber, F., and Schaffner, W. (1985) *Nature* 315, 75-77.

39. Wildeman, A.G., Sassone-Corsi, P., Grundstrom, T., Zenke, M., and Chambon, P. (1984) *EMBO J.* 3, 3129-3133.

40. Wildeman, A.G., Zenke, M., Schatz, C., Wintzerith, M., Grundstrom, T., Matthes, H., Takahashi, K., and Chambon, P. (1986) *Mol. Cell. Biol.* 6, 2098-2105.

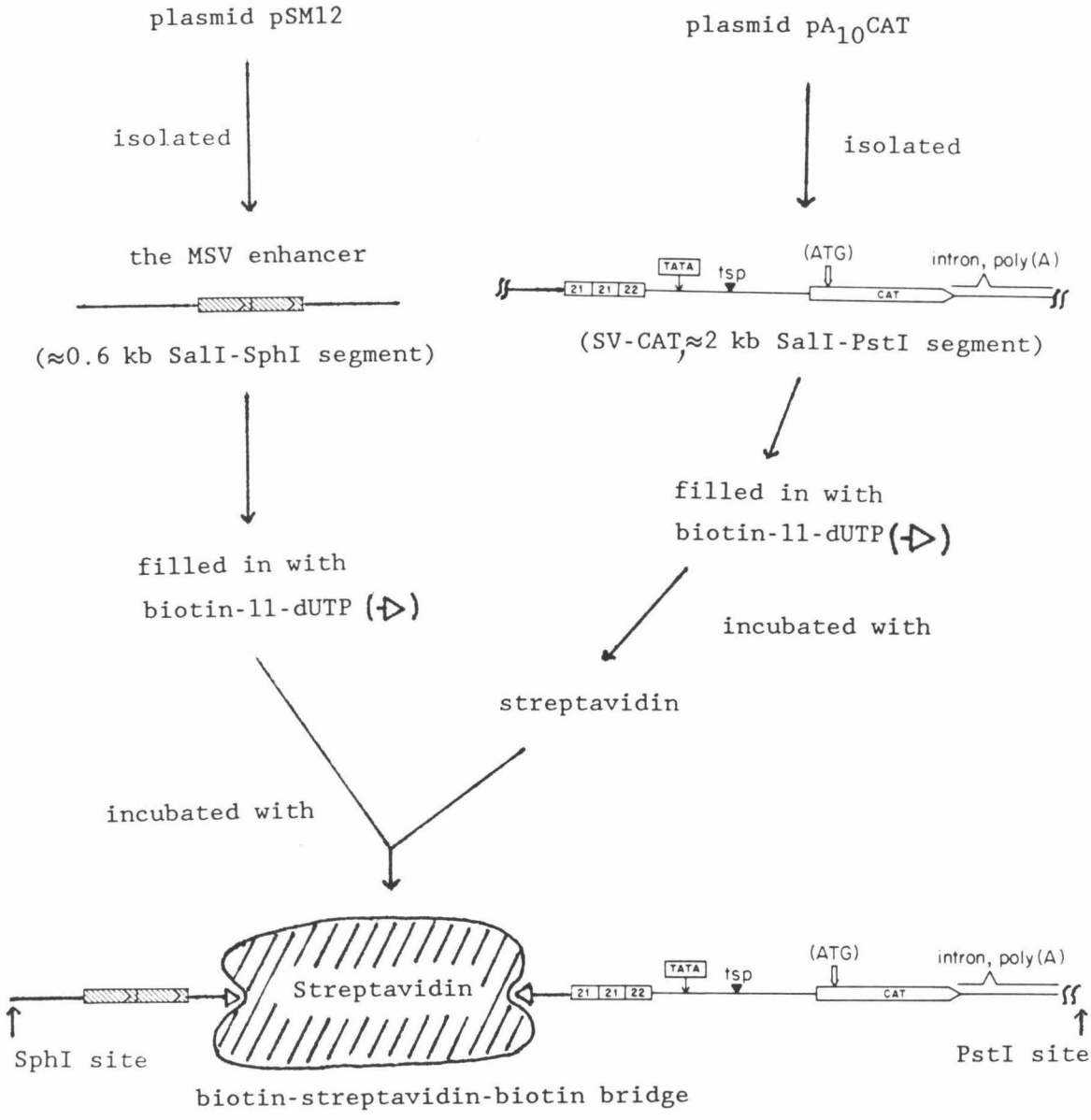


Figure 1. Schematic diagram of the construction of the composite enhancer/transcription unit with a biotin-streptavidin-biotin bridge.

PROPOSITION 5

The use of proteolytic digestion and a new chromatography technique to probe the functional domains and the conformation of the human nerve growth factor receptor.

ABSTRACT

Nerve growth factor (NGF) exerts its effects on NGF-dependent neurons by interacting with specific, high-affinity receptors located on the neuronal cell surface. A model for binding of NGF to its receptor has been proposed previously (Buxser et al., 1985). A high affinity NGF/NGF-receptor complex probably results from a dimeric NGF binding to a dimeric NGF-receptor, so that each monomer of NGF associates with a separate receptor molecule, whereas a low affinity complex results from a dimeric NGF binding to a single receptor molecule (a receptor monomer). However, the molecular basis for the interaction between NGF and its receptor is not understood. I therefore propose to study the functional domains (NGF-binding sites) and their conformation (dimeric form) of the NGF receptor using a proteolytic dissection approach and a new chromatography technique. If successful, this proteolytic dissection method may be useful for structure determination for other growth factor receptors, and the proposed chromatography technique may prove to be a very sensitive and rapid method for separating rare proteins under non-denaturing condition.

INTRODUCTION

During development of the vertebrate nervous system, nerve growth factor (NGF) plays an essential role in the survival and maintenance of sympathetic and sensory neurons (Levi-Montalcini and Angeletti, 1968). NGF was the first and is the only neurotrophic growth factor identified so far that has been shown to be responsible for neuronal survival *in vivo*. NGF exerts its effects on NGF-dependent neurons by interacting with specific, high-affinity receptors located on the neuronal cell surface. The human NGF receptor has been purified (Puma et al., 1983, Grob et al., 1985), its cDNA clones have been isolated (Johnson et al., 1986), and the structure of the gene determined (Chao et al., 1986). However, the molecular basis for the mechanism of NGF action is not understood. The deduced amino acid sequence of the human NGF receptor (Johnson et al., 1986) reveals several features that distinguish it from other receptors for growth factors. First, the NGF receptor has no sequence homology with the family of genes encoding either serine/threonine or tyrosine specific protein kinases. It is thus unlike a variety of other growth factors including epidermal growth factor (EGF) (Ushiro and Cohen, 1980), platelet-derived growth factor (PDGF) (Ek et al., 1982), colony-stimulating factor (CSF-1) (Sherr et al., 1985), and fibroblast growth factor (FGF) (Thomas et al., 1984). Second, there is no sequence homology between NGF and insulin receptors, whereas sequence homology between NGF and insulin has been reported (Frazier et al., 1972). Third, although NGF stimulates serine/threonine phosphorylation of specific nuclear and cytoskeletally associated proteins (Yu et al., 1980; Landreth and Rieser, 1985), the protein kinase activities

responsible for these phosphorylation reactions apparently may not be intrinsic to the receptor since the NGF receptor lacks significant sequence homology with known serine/threonine protein kinases. Therefore, understanding the mechanism of the interaction of NGF with its receptor is of fundamental importance in the NGF-receptor signal transduction mechanism and in developing an overall picture of the development of NGF-dependent neurons.

Characterization of the cell surface NGF-receptor by kinetic and equilibrium binding parameters has revealed two apparent affinity states (Sutter et al., 1979; Landreth and Shooter, 1980; Schechter and Bothwell, 1981; Buxser et al., 1983). The first NGF-receptor state has a high affinity for ^{125}I-NGF (dissociation constant $K_d \approx 3 \times 10^{-10}$M), and the second state has a lower affinity for ^{125}I-NGF ($K_d \approx 3 \times 10^{-9}$M). Photoaffinity cross-linking of ^{125}I-NGF experiments indicated that two NGF-binding proteins existed in membrane preparations (Buxser et al., 1983; Grob et al., 1983). Furthermore, analysis of highly purified NGF-receptor revealed two proteins with molecular weight of 80- and 200-kilodalton (kDa) (Puma et al., 1983). A model for binding of NGF to its receptor has been proposed according to the dimeric structure of NGF and binding properties of NGF to its receptor (Buxser et al., 1985). The biologically active form of NGF is composed of two identical 13-kDa peptides which are noncovalently associated to form a 26-kDa molecule (Greene and Shooter, 1980). The dimeric structure of NGF suggests that each 13-kDa monomer may be able to bind to one NGF-receptor. As shown in Figure 1 (page 149), when a dimeric NGF binds to a single 80-kDa receptor molecule it forms a low affinity complex. When each monomer of

NGF associates with a separate receptor molecule it will form a high affinity complex that is a dimeric receptor structure bridged by the dimeric NGF (200-kDa complex). Simultaneous dissociation of NGF from two receptors is significantly less probable than dissociation from a single receptor. Apparently, with two receptors there is a slower dissociation rate and higher affinity. This model is supported by the evidence that the NGF-receptor exists, at least partially, as a disulfide-linked dimer (Grob et al., 1985).

Several years ago, Geisler and Weber (1977) developed a method for directly assessing the role of the amino terminus of *lac* repressor in operator binding. They found that, in 1.0 M Tris-HCl, pH 7.5/ 30% (V/V) glycerol, tryptic cleavage is limited to only two sites in the *lac* repressor arginine-51 and lysine-59. Cleavage under these conditions yielded the tetrameric trypsin-resistant core and approximately equal amounts of two amino-terminal "headpieces" which could bind specifically to the *lac* operator (Ogata and Gilbert, 1978). Similarly, using papain digestion of the λ phage repressor, Pabo et al. (1979) demonstrate that the amino-terminal domain binds DNA, and the carboxyl-terminal domain oligomerizes. Moreover, Simmons (1988) has identified very recently the regions of the simian virus 40 (SV40) large T antigen that make strong, intermediate, and weak bindings to the viral origin of DNA synthesis after digestion with various concentrations of Pronase E, a nonspecific protease. It should be possible therefore to apply the proteolytic dissection technique in a similar way to identify the functional domains of the human NGF receptor.

SPECIFIC AIMS

The goal of this proposal is to design an experimental approach to identify the functional domains of the human NGF receptor, and further to determine their conformation under physiological conditions. This should lead, in the long term, to an understanding of the structure of the NGF-binding sites in the receptor protein, and ultimately to an elucidation of the mechanism of NGF action.

METHODS AND PROCEDURES TO BE USED IN PROPOSED PROJECT

I propose to use the proteolytic dissection technique to identify the functional domains of the human NGF-receptor. First, I will determine the proper conditions for digestion of the NGF/NGF-receptor complexes with Pronase E. Affinity chromatography-purified NGF-receptor (Puma et al., 1983) will be incubated with ^{125}I-labeled-NGF and then digested with various concentrations of Pronase E. The size of the digested peptides will be analyzed by electrophoresis in sodium dodecyl sulfate (SDS) 15% polyacrylamide gels (Laemmli, 1970) in the presence of dithiothreitol. Alternatively, a time course of Pronase E digestion of the NGF/NGF-receptor complexes can be performed and monitored by the same gel electrophoresis. Second, purified NGF will be coupled to CNBr-activated Sepharose 4B (Pharmacia) for 16 hours at 4 ^{o}C in 0.5M NaCl, 0.1M borate, pH 8.5 buffer as described (Puma et al., 1983). After coupling is completed, the gel will be washed and equilibrated in equilibration buffer (10mM Na-phosphate, pH 7.2, 136mM NaCl, 5mM KCl, 1.2% octylglucoside and phosphatidylcholine). This column will be used as an affinity matrix for NGF receptors. Purified NGF receptors will be

passed over the column and incubated for 1 hour at 22 °C. The column will then be extensively washed with equilibration buffer. Subsequently, Pronase E (in equilibration buffer) at the proper concentration as determined above will be passed through the column and incubate for a certain time as determined in the first step. Third, after an extensive wash step, any NGF-receptor fragments still bound to the column will be eluted with buffer of 0.15M NaCl, pH 11.0, containing 1.2% phosphatidyl-choline. The origin of the eluted fragments (i.e. the NGF-binding domains) on NGF receptor will be determined by automated Edman degradation and amino acid analysis as described (Hunkapiller et al., 1984). As a control, nonspecific binding of the NGF receptor to Sepharose-4B will be determined by parallel incubation in the column not containing any NGF. In addition, other "amino acid-specific" proteases may be used in the digestions as described above. For instance, trypsin which cleaves after arginine and lysine, chymotrypsin which cleaves after phenylalanine, endoproteinase Lys-C which cleaves lysine and endoproteinase Arg-C which cleaves arginine specifically (Levy et al., 1970), and endonuclease Glu-C (V8 protease) which hydrolyzes peptides specifically at the carboxyl-terminal of glutamate residues (Houmard and Drapeau, 1972) can all be used.

Since the high affinity NGF receptor exists as a disulfide-linked dimer, it can be determined whether or not the NGF-binding domains still retain the dimeric conformation. In order to rapidly determine molecular weights of "native", undissociated proteolyzed protein fragments, I therefore propose a new gel permeation chromatography. The basic idea of this method is to perform gel filtration on a glass plate

as thin-layer chromatography rather than carrying out column chromatography (see below). Several advantages of this proposed method justify the attempt to develop it. First, the separated peptides can be transferred from the gel layer to nitrocellulose filter membrane (see below), therefore, they can be detected by not only the conventional gel-staining methods such as coomassie blue and silver staining (Merril et al., 1979), but also with radioactively labeled antibodies specific to sample proteins. By contrast, in column chromatography or gel filtration HPLC (high performance liquid chromatography), the separated proteins usually can be detected only by the ultraviolet (VU) absorbance. Thus, one can use more sensitive detecting methods in this proposed chromatography technique than in the conventional column chromatography or gel filtration HPLC. Second, many experimental samples and molecular weight standards can be easily applied on a single gel-plate and compared because it is a two dimensional gel filtration (see below). By contrast, only one unknown sample, with or without molecular weight standards, can be applied on a column chromatography or gel filtration HPLC. Third, the conditions of filtration should be very constant from sample to sample in this proposed chromatography since all of the samples are present simultaneously on the same gel-plate. Fourth, unlike column chroma-tography, there is no need to determine the void volume of the gel-plate beforehand in this proposed chromatography. Moreover, the filtration data by the proposed method may provide the elution condition for column chromatography or gel filtration HPLC. Fifth, although estimation of molecular weights of "native", undissociated proteins may be achieved by disc gel electrophoresis (Hendrick and

Smith, 1968; Kempe et al., 1974), it is generally more cumbersome than gel filtration.

The detailed procedures for this proposed chromatography are described as the following. The preparation of Sephadex G-100 (superfine grade) will be the same as that for column chromatography. After swelling, the gel particles will be allowed to settle and the supernatant removed by decantation. The gel slurry will be poured onto a glass plate horizontally, spread out evenly on the plate, and allowed to settle undisturbed for 2 or more hours in a humid chamber until it forms a thin-layer of gel (still moist) on the glass plate. Then, the gel-plate will be set up as Figure 2A (page 150). The buffer has to be run over the plate by descending chromatography. The bridges carrying the eluant to and from the plate have to be thick filter paper (e.g. Whatman 3 MM) in order to have sufficient capacity. Note that the filter paper should be covered with plastic wrap to avoid evaporation. The flow rate through the gel can be regulated by the angle of the plate to the horizontal.

Before the run is started, the plate has to be equilibrated by allowing the sovent to flow through the gel for several hours. After equilibration, the sample (2-5 μl) will be applied through a micropipette as a round spot (ca. 2-4 mm in diameter). Sample application should be carried out with the plate horizontal and with care not to disturb the gel layer. The flow rate over the plate can be measured by the use of a colored reference substance, e.g. blue dextran. After completion of the run, as shown in Figure 2B (page 150), the separated peptides will be transferred from the gel layer by absorption into a

superimposed sheet of nitrocellulose filter membrane as the Southern transfer procedure for DNA samples (Southern, 1975). This step should be performed with the plate horizontal and with care not to disturb the gel layer. It is important that air bubbles not be trapped between the membrane and the gel layer with this procedure. After about 6-12 hours of transfer, the nitrocellulose membrane will be removed and then air dried and stained according to conventional methods, e.g. coomassie blue or silver staining (Merril et al., 1979). Alternatively, it can be probed with radioactively labeled antibodies specific to sample proteins after the membrane has pre-hybridized with the blocking solution to eliminate the nonspecific binding of antibodies.

REFERENCES

1. Buxser, S.E., Kelleher, D.J., Watson, L., Puma, P., and Johnson, G.L. (1983) *J. Biol. Chem.* 258, 3741-3749.

2. Buxser, S.E., Watson, L., and Johnson, G.L. (1983) *J. Cell. Biochem.* 22, 219-233.

3. Buxser, S., Puma, P., and Johnson, G.L. (1985) *J. Biol. Chem.* 260, 1917-1926.

4. Chao, M.V., Bothwell, M.A., Ross, A.H., Koprowski, H., Lanahan, A., Buck, C.R., and Sehgal, A. (1986) *Science* 232, 418-421.

5. Ek, B., Westermark, B., Wasteson, A., and Heldin, C.-H. (1982) *Nature* 295, 419-420.

6. Frazier, W.A., Angeletti, R.H., and Bradshaw, R.A. (1972) *Science* 176, 482-488.

7. Geisler, N., and Weber, K. (1977) *Biochemistry* 16, 938-943.

8. Grob, P.M., Berlot, C.H., and Bothwell, M.A. (1983) *Proc. Natl. Acad. Sci. USA* 80, 6819-6823.

9. Grob, P.M., Ross, A.H., Koprowski, H., and Bothwell, M.A. (1985) *J. Biol. Chem.* 260, 8044-8049.

10. Hedrick, J.L. and Smith, A.J. (1968) *Arch. Biochem. Biophys.* 126, 155-164.

11. Houmard, J. and Drapeau, G.R. (1972) *Proc. Natl. Acad. Sci. USA* 69, 3506-3509.

12. Greene, L.A., and Shooter, E.M. (1980) *Annu. Rev. Neurosci.* 3, 353-402.

13. Hunkapiller, M., Kent, S., Caruthers, M., Dreyer, W., Firca, J., Giffin, C., Horvath, S., Hunkapiller, T., Tempst, P., and Hood, L.

(1984) *Nature* 310, 105-111.

14. Johnson, D., Lanahan, A., Buck, C.R., Sehgal, A., Morgan, C., Mercer, E., Bothwell, M., and Chao, M. (1986) *Cell* 47, 545-554.
15. Kempe, T.D., Gee, D.M., Hathaway, G.M., and Noltmann, E.A. (1974) *J. Biol. Chem.* 249, 4625-4633.
16. Laemmli, U.K. (1970) *Nature* (London) 227, 680-685.
17. Landreth, G.E., and Shooter, E.M. (1980) *Proc. Natl. Acad. Sci. USA* 77, 4751-4755.
18. Landreth, G.E., and Rieser, G.D. (1985) *J. Cell Biol.* 100, 677-683.
19. Levi-Montalcini, R., and Angeletti, P.U. (1968) *Physiol. Rev.* 48, 534-569.
20. Levy, M., Fishman, L., and Schenkein, I. (1970) *Methods Enzymol.* 19, 672-681.
21. Merril, C.R., Switzer, R.C., and Van Keuren, M.L. (1979) *Proc. Natl. Acad. Sci. USA* 76, 4335-4339.
22. Ogata, R. and Gilbert, W. (1978) *Proc. Natl. Acad. Sci. USA* 75, 5851-5854.
23. Pabo, C.O., Sauer, R.T., Sturtevant, J.M., and Ptashne, M. (1979) *Proc. Natl. Acad. Sci. USA* 76, 1608-1612.
24. Puma, P., Buxser, S.E., Watson, L., Kelleher, D.J., and Johnson, G.L. (1983) *J. Biol. Chem.* 258, 3370-3375.
25. Schechter, A.L., and Bothwell, M.A. (1981) *Cell* 24, 867-874.
26. Sherr, C.J., Rettenmier, C.W., Sacca, R., Roussel, M.F., Look, A. T., and Stanley, E.R. (1985) *Cell* 41, 665-676.
27. Simmons, D.T. (1988) *Proc. Natl. Acad. Sci. USA* 85, 2086-2090.
28. Southern, E.M. (1975) *J. Mol. Biol.* 98, 503-517.

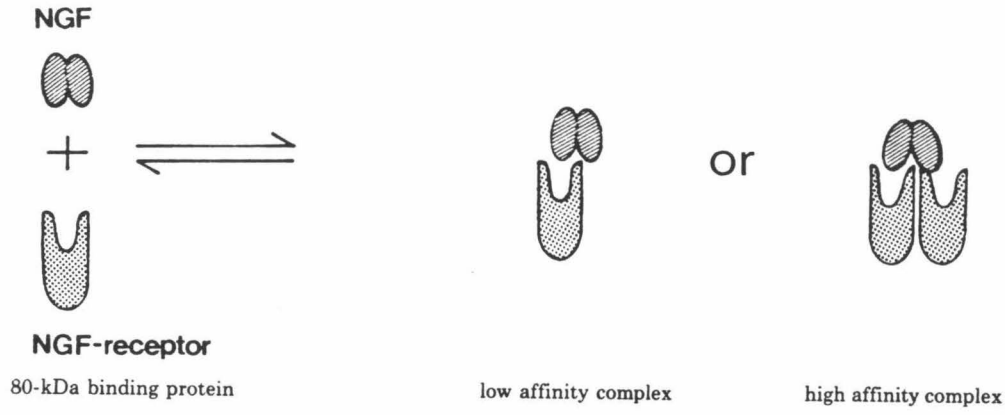
29. Sutter, A., Riopelle, R.J., Harris-Warrick, R.M., and Shooter, E. M. (1979) *J. Biol. Chem.* 254, 5972-5982.

30. Thomas, K.A., Rios-Candelore, M., and Fitzpatrick, S. (1984) *Proc. Natl. Acad. Sci. USA* 81, 357-361.

31. Ushiro, H., and Cohen, S. (1980) *J. Biol. Chem.* 255, 8363-8365.

32. Yu, M.W., Tolson, N.W., and Guroff, G. (1980) *J. Biol. Chem.* 255, 10481-10492.

(A)



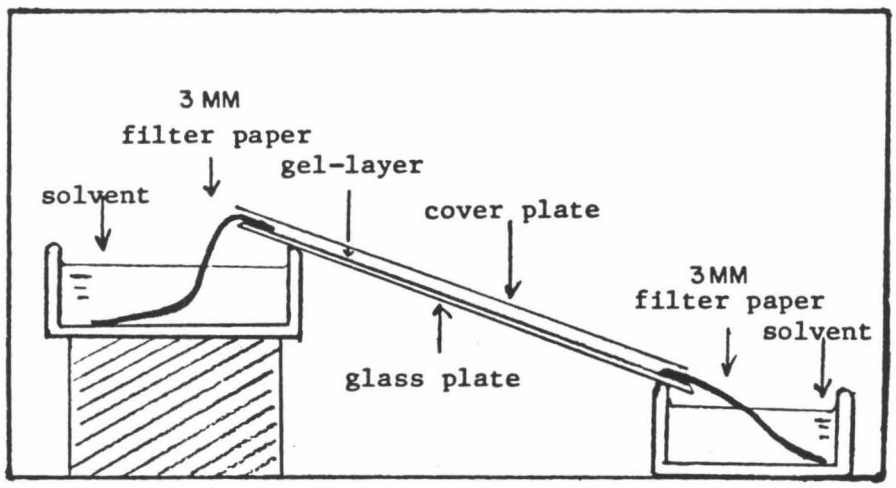
(B)

$$\text{NGF} + \text{80-kDa binding protein} \rightleftharpoons \text{low affinity complex}$$

$$\text{Low affinity complex} + \text{80-kDa binding protein} \rightleftharpoons \text{high affinity complex}$$

Figure 1. A model for binding of NGF to NGF-receptor to form low and high affinity complexes.

(A)



(B)

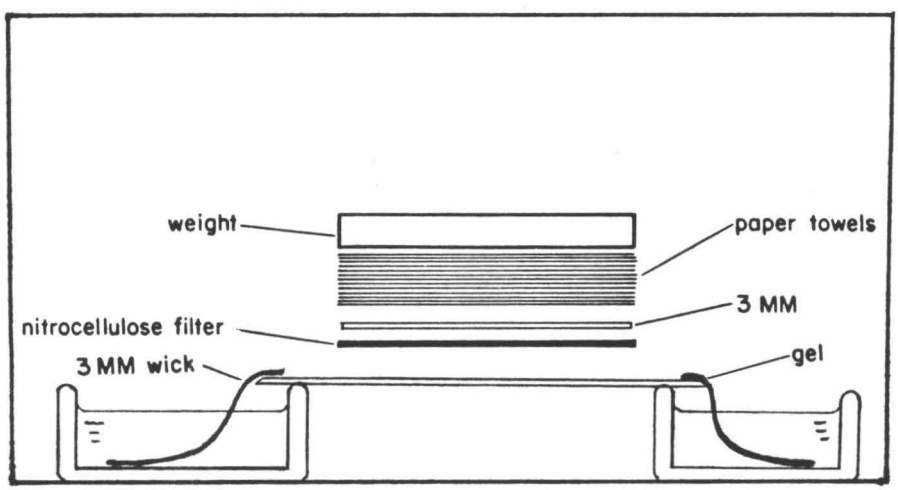


Figure 2. A set-up for the proposed gel permeation chromatography (A), and the "Southern"-like protein transfer (B).