4. E. R. Kandel, *Behavioral Biology of Aplysia* (Freeman, New York, 1979).
5. W. T. Frazier, E. R. Kandel, I. Kupferman, R. Waziri, R. E. Coggeshall, *J. Neurophysiol.* **30**, 1288 (1967).
6. R. E. Coggeshall, B. A. Yaksta, F. J. Swartz, *Chromosoma* **32**, 205 (1971).
7. D. Gardner, *Science* **173**, 550 (1971).
8. D. L. Wilson, *J. Gen. Physiol.* **57**, 26 (1971); Y. P. Loh, Y. Sarne, M. P. Daniels, H. Gainer, *J. Neurochem.* **29**, 135 (1977).
9. R. H. Scheller, J. F. Jackson, L. B. McAllister, J. H. Schwartz, E. R. Kandel, R. Axel, *Cell* **28**, 707 (1982).
10. R. H. Scheller, J. F. Jackson, L. B. McAllister, B. S. Rothman, E. Mayeri, R. Axel, *ibid.* **32**, 7 (1983); J. R. Nambu, R. Taussig, A. C. Mahon, R. H. Scheller, *ibid.* **35**, 47 (1983); R. Taussig, R. R. Kaldany, R. H. Scheller, *Proc. Natl. Acad. Sci. U.S.A.*, in press; A. C. Mahon and R. H. Scheller, in preparation.
11. R. Taussig, M. R. Picciotto, R. H. Scheller, in *Molecular Biology of Development*, E. H. Davidson and R. I. Firtel, Eds. (Liss, New York, in press), vol. 19.
12. L. B. McAllister, R. H. Scheller, E. R. Kandel, R. Axel, *Science* **222**, 800 (1983).
13. E. Herbert and M. Uhler, *Cell* **30**, 1 (1982); E. Herbert *et al.*, *Neurosci. Commentaries* **1**, 16 (1981).
14. F. Strumwasser, L. K. Kaczmarek, A. Y. Chiu, E. Heller, K. R. Jennings, D. P. Viele, in *Peptides: Integrators of Cell and Tissue Function*, F. E. Bloom, Ed. (Raven, New York, 1980), p. 197.
15. T. Kreiner, J. Rothbard, G. K. Schoolnick, R. H. Scheller, *J. Neurosci.*, in press.
16. C. H. Price, D. J. McAdoo, W. Farr, R. Okuda, *J. Neurobiol.* **10**, 551 (1979); C. H. Price and D. J. McAdoo, *Brain Res.* **219**, 307 (1981).
17. E. Giller and J. H. Schwartz, *J. Neurophysiol.* **34**, 93 (1971).
18. E. Mayeri, P. Brownell, W. D. Branton, *ibid.* **42**, 1165 and 1185 (1979).
19. B. S. Rothman, E. Mayeri, R. O. Brown, P. M. Yuan, J. E. Shively, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 5753 (1983).
20. M. Schaefer, P. Ruben, S. H. Thompson, R. H. Scheller, unpublished data.
21. P. H. Brownell and M. Schaefer, *Soc. Neurosci. Abstr.* **8**, 736 (1982).
22. S. Mackey and T. J. Carew, *J. Neurosci.* **3**, 1469 (1983).
23. D. K. Stuart and F. J. Strumwasser, *J. Neurophysiol.* **48**, 499 (1980).
24. H. R. Morris, M. Panico, A. Karplus, P. E. Lloyde, B. Riniker, *Nature (London)* **300**, 643 (1982); P. E. Lloyd, I. Kupfermann, K. R. Weiss, *Proc. Natl. Acad. Sci. U.S.A.*, in press.
25. We thank our many collaborators, including C. Evans, G. Makk, J. Rothbard, and G. Schoolnik. Supported by grants to R.H.S. from the National Institutes of Health and the March of Dimes Foundation. R.H.S. is a McKnight Foundation scholar and an Alfred P. Sloan research fellow.

Control of Neuronal Gene Expression

J. Gregor Sutcliffe, Robert J. Milner
Joel M. Gottesfeld, Wanda Reynolds

Cellular phenotype is the direct consequence of the concentration of all the proteins expressed by the cell. Although synthesis and accumulation of proteins are sensitive to control at many steps, it is generally accepted that the control of transcriptional initiation of messenger RNA (mRNA) is the most consequential of all these steps in determining phenotype. In higher eukaryotes, the ability of cells to form organs or tissues with their own specific functions is the result of these cells producing unique sets of proteins during the course of the developmental program. Because all cell types are believed to contain the same DNA [with the known exception of immunoglobulin (Ig) genes in producer cells], events must occur during development that alter the transcriptional states of sets of genes (epigenesis). For genes that are exclusively expressed in one organ, transcriptional initiation is the all-or-none switch. Such genes are likely to give us the clearest insights as to how the differentiation process operates.

Summary. Some 30,000 genes are expressed exclusively in the rat brain, many of which contain a genetic element called an identifier sequence located in at least one of their introns. The identifier sequences are transcribed by RNA polymerase III exclusively in neurons to produce two RNA species, BC1 and BC2, of 160 and 100 to 110 nucleotides. This transcriptional event may define regions of chromatin that contain neuronal-specific genes and may poise these genes for transcription by polymerase II by rendering the gene promoters accessible to soluble *trans*-acting molecules.

Transcriptional initiation in prokaryotes is well understood. Transcription begins at promoters, specific DNA sequences located 5′ to the start of RNA polymerization. These sequences respond to a particular combination of RNA polymerase and so-called sigma or specificity factors. In some cases, promoter selection is influenced by other proteins, such as catabolite activator protein (CAP). In simple prokaryotes such as *Escherichia coli*, there is only one known sigma factor. For bacteria (such as *Bacillus subtilis*) that exhibit different life forms during processes such as sporulation, the available sigma factors that drive particular genes vary according to life stage. Variable gene regulation in prokaryotic organisms can be simply divided into two classes: positive control, which involves sigma factors and other positive activators (such as CAP) available only in response to certain external conditions, and negative control, which involves repressor molecules that inactivate genes by interacting with their promoters to prevent RNA polymerase from initiating gene expression. At any particular time, all genes are equally available to interact with each positive or negative *trans*-regulatory molecule.

Eukaryotic organisms are obviously more complicated. What formally separates higher eukaryotes from prokaryotes is that the cells of higher eukaryotes must first differentiate into organs, each composed of one or more specific cell types, and then those organs must be able to respond to ambient stimuli reflecting the physiological state of the whole organism. Batteries of genes must be activated in response to certain stimuli, and the required responses to the same stimulus may be different for the cells of different tissues. Actual development of higher animals involves several distinct stages for each differentiating cell type. The cumulative result of these processes is that, in developing and adult organisms, different genes are available for activation in different tissues at different stages. This implies, a priori, two distinct levels of gene control, one related to the differentiation state (both temporal and type) of the particular cell and

Drs. Sutcliffe and Gottesfeld are associate members and Dr. Reynolds is a research associate in the Department of Molecular Biology and Dr. Milner is an assistant member in the Department of Basic and Clinical Research, MB4, Scripps Clinic and Research Foundation, 10666 North Torrey Pines Road, La Jolla, California 92037.

the other related to the response of particular genes to ambient signals. The phenotypic manifestations of the two levels of necessary gene regulation are known to developmental biologists and are classically referred to as effects of cell determination and cell position. We should expect that an individual gene might contain sequences enabling activation of that gene in the proper organ according to a cell's history or lineage and also sequences that can respond to ambient signals. Therefore, whereas the important *cis*-acting sequences governing the control of prokaryotic genes cluster in small regions very near the initiation site for gene transcription, one expects that more extensive regions may be involved in *cis*-control of eukaryotic transcription. Many recent studies have shown that certain regulators may act at a considerable nucleotide distance from eukaryotic promoters.

Models of eukaryotic gene control derive from the early work of Britten and Davidson, who built on the models of Jacob and Monod for prokaryotic gene control. It was realized at an early stage that batteries of genes that exhibit common regulation must contain common control sequences, but this now obvious realization was not experimentally accessible until recently. Britten and Davidson (*1*) originally proposed that moderately repetitive DNA sequences acted as control elements regulated by tissue-specific RNA or protein molecules. Although a few of the specific aspects of these early models have not been borne out, the models serve to provide a generally accepted formal view of eukaryotic gene control. Evidence for common *cis*-acting regulatory sequences and *trans*-acting regulatory molecules now exists for several systems.

Promoters Respond to Ambient Signals

Some *cis*-acting sequences are located 5′ to genes in regions close to RNA polymerase II initiation sites (polymerase II is the enzyme responsible for the production of mRNA in eukaryotes). The heat shock genes of *Drosophila* are activated in response to thermal stress. Upstream from each heat shock gene is a common promoter element, which if placed 5′ to other genes causes these genes to become heat-inducible (*2*). Deoxyribonuclease protection (footprinting) studies have provided evidence for a specific factor that interacts with these control elements (*3*). Several yeast amino acid biosynthetic enzymes are coordinately expressed after amino acid starvation (*4*). Multiple copies of an 8-bp (base pair) sequence are located upstream from the genes coding for these enzymes. This sequence is required for the starvation response, and a gene for a putative regulatory protein that acts in *trans* has been identified (*5*). Efficient transcription in vitro of the SV40 (Simian virus 40) early gene region requires the presence of the tandem 21-bp repeats located 70 to 100 bp upstream from the RNA start site. HeLa extracts contain two components, Sp1 and Sp2, in addition to polymerase II, that are required for transcription of the SV40 early genes. Sp2 is a general transcription factor, but Sp1 is specific for the SV40 early promoter. Footprinting studies with Sp1 show that it binds to the 21-bp repeats (*6*). The expression of mouse mammary tumor virus (MMTV) is regulated by ambient concentrations of glucocorticoids. Glucocorticoid receptor protein binds to a region of DNA upstream from the RNA start site in the MMTV long terminal repeat (LTR) as well as a few sites internal to the MMTV genome. Insertion of the receptor-binding sequence upstream from a heterologous gene causes the promoter of that gene to come under hormonal control (*7*). Thus, DNA sequences within a few hundred nucleotides 5′ to genes are involved in the regulation of polymerase II transcription. The factors affecting these sites in *trans* distinguish between promoters, and some promoters become active in response to particular external stimuli. Although these systems are more complex to analyze experimentally than bacterial promoters, at the formal level the events that occur at eukaryotic promoters seem analogous to those involved in the positive control of prokaryotic promoters. Many of these systems seem to be consistent with those which monitor ambient signals, that is, signals of cell position.

Enhancer Elements as Distant Positive Regulators

Another class of *cis*-acting positive regulatory elements are the enhancers. The 72-bp tandem repeats of SV40 stimulate transcription of the T antigen gene and also a variety of cellular genes when inserted anywhere within several thousand base pairs upstream or downstream from the RNA start site; they function in either orientation relative to the gene (*8, 9*). Deletion analyses suggest that the 3′ region of the SV40 enhancer contributes to function but is dispensable; the 5′ portion is absolutely essential (*9–11*). Functionally analogous enhancer elements are known from work on other viruses; these include polyoma virus, bovine papilloma virus, and retrovirus LTR's (*12–14*). There is also an enhancer element in an intron of active mouse and human Ig genes. Activation of this element results from the juxtaposition of the enhancer with the rearranged V segment (*15*). Although there are no extensive homologies found by sequence comparisons of the various enhancer elements, Weiher and his colleagues (*11*) have proposed a consensus enhancer core sequence of GTGG$^{TTT}_{AAA}$G (G, guanine; T, thymine; A, adenine), which is found in all described enhancers and is the site of mutations that reduce the efficiency of the SV40 enhancer. Another consistent feature of enhancer sequences is that they often are near regions of alternating purines and pyrimidines, a structure capable of forming Z DNA (*16*).

It is not merely the potential Z-like structure of enhancer DNA that gives it its stimulatory effect on transcription. The enhancers of SV40 and those of Moloney murine leukemia and sarcoma virus LTR's are interchangeable. However, whereas the SV40 enhancer is five times more potent than the mouse retroviral enhancer in monkey cells, the mouse enhancer is twice as potent as the SV40 sequence in mouse L cells (*12, 14, 17*). Consequently, enhancer activity is not solely a result of inherent DNA structure, and there must, therefore, be soluble factors (that is, proteins) that differ between species and determine this preferential, but not absolute, effect. Competition transfection studies provide evidence for the limiting quantities of these enhancer recognition factors (*18*). When the SV40 enhancer is replaced by the Ig enhancer, SV40 T antigen is expressed in transfected lymphoid cell lines but not in other mouse or HeLa cells, suggesting an interaction of the Ig enhancer with proteins present in only a specific cell type (*19*). Thus, it seems that some enhancer elements function in all cell types (although perhaps better in some cells than others), while other enhancer elements show more cellular specificity. The proteins acting on these enhancers are regulated in an unknown manner. The mechanism of their action, that is, how action at one site is communicated to a polymerase II promoter several kilobases away, is also unknown, although in a general way it is thought to involve the alteration of DNA or chromatin structure. From what is known thus far, it is not possible to decide whether enhancer sequences are mark-

Table 1. Characteristics of brain mRNA from hybridization data (*26*). For these values we have assumed that the 26 percent of brain mRNA that is very rare (less than 0.01 percent and not detected in blotting hybridization experiments) is mostly brain-specific, but other reasonable assumptions generate similar values. The average nucleotide lengths in parentheses are those actually measured; the nonparenthetical values are those calculated. The number average takes into account the amount of each mRNA as well as its length.

mRNA class	RNA mass (%)	Average nucleotide length	Number average nucleotide length
I: Same concentration in all tissues	18	(1780) 2680	1250
II: Variable concentration in different tisues	26	(2350) 2270	1870
III: Brain-specific	56	(3600) 4960	2640

ers of cell lineage or cell position. We will develop an argument that suggests that some features of enhancer sequences are consistent with lineage-specific markers. Some of these features are common with features of certain RNA polymerase III promoters.

Elements of Cell Lineage

A third category of sequences which may be involved in gene control are those defined by their recurrence in coordinately controlled transcripts rather than by explicit functional demonstration of their critical role in transcription. Thus the implication of these sequences is circumstantial but, in some cases, provocative and compelling. A common sequence is found in the mRNA's of several coordinately controlled *Dictystelium* mRNA's (*20*). A different common sequence, called an identifier (ID) sequence, is located in an intron of most or all adult neuronal-specific mRNA precursors (*21–23*). Another sequence is found in the noncoding region at the 3′ ends of several mRNA's activated in tumor cell lines (*24*). Finally, the transcription of repetitive sequences of sea urchins is different at different developmental stages (*25*).

Neuronal Lineage Elements

The studies that led to the definition of a neuronal-specific pre-mRNA sequence were motivated by an intent to define neuronal-specific proteins. Because mRNA's and proteins are likely to be equivalent sets for any organ, and because to understand an organ we need to know what its specific proteins are and how they function, we constructed complementary DNA (cDNA) clones of rat brain mRNA's (*26*). We analyzed these clones by RNA (Northern) blot analysis, comparing the hybridization of some 190 individual nick-translated cDNA clones to the polyadenylated [poly(A)$^+$] RNA from brain, liver, and kidney, thus learning the size, tissue distribution, and relative abundances of a cross section of brain mRNA's (summarized in Table 1). We found that up to 56 percent of the brain mRNA mass (some 30,000 distinct, mostly rare mRNA species) was exclusively expressed in the brain (*26*). These mRNA's have the surprising property that they average 5000 nucleotides in length, whereas other mRNA's average about 2000 nucleotides in length. Another 26 percent of the mRNA mass represents species shared by the three test tissues; however, the amount in each cell type is highly variable, indicating that their genes are regulated in a tissue-specific manner. The remaining 18 percent of the molecules are equally expressed in each of the three test tissues.

We have utilized the nucleotide sequences of some of these cDNA clones to deduce the amino acid sequence of the proteins encoded by the brain-specific mRNA's. We then made synthetic peptides corresponding to short regions of the proteins and used antisera to the synthetic peptides to describe the anatomical distribution and molecular characteristics of rare brain-specific proteins, including the precursor for a putative neurotransmitter 1B236, whose synthesis we will discuss below [(*27*); these experiments are discussed in this issue by L. F. Reichardt *et al.*, page 1294].

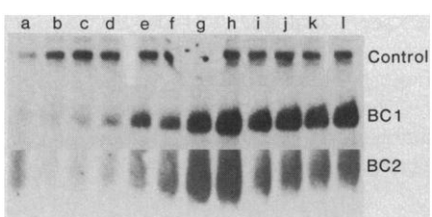

Fig. 1. Developmental time course for the accumulation of BC1 and BC2 RNA's. Poly(A)$^+$ RNA was isolated from the brain of pre- and postnatal rats, and equal masses of tissue at each stage was analyzed by RNA blotting with ^{32}P-labeled ID probes. (Lanes a, b, c, and d) embryonic days 14, 16, 18, and 20; (lanes e, f, g, h, i, j, k, and l) postnatal days 1, 5, 10, 15, 20, 25, 30, and adult. As a control, the neuronal-specific mRNA 1A1075 is shown. Day 14 is low for this probe because RNA was prepared from whole head, which contains nonneuronal tissue; hence neuronal RNA is diluted. The BC1 and BC2 strips are different exposures of the same blot; the control is a parallel blot that has been cut between lanes d and e and k and l to remove inadvertent empty lanes. Probes and blots were prepared as described (*22*).

Three percent of our cDNA clones hybridized to a brain-specific 160-nucleotide poly(A)$^+$ RNA species, even though the clones had much larger (561 to 1225 bp) cDNA inserts (*21*). The small RNA (called BC1 for brain cytoplasmic) is absent from adrenal, spleen, testes, heart, muscle, intestine, lung, thymus, liver, and kidney. It is present in brain, pituitary, and solar plexus (a peripheral nervous tissue) (*22*). We believe that BC1 is a neuronal marker because it is absent from glial tumor lines and because the relative distribution of BC1 is inversely related to the glial distribution in the brain.

BC1 RNA is found almost exclusively in the cytoplasm (about 2000 molecules per cell) and migrates together with 5.8*S* RNA. In highly purified brain poly(A)$^+$ RNA preparations, BC1 is clearly visible on stained gels. A second RNA species, BC2, migrates as a diffuse smear of 100 to 110 nucleotides, hybridizes with the same probes, and seems to share all other properties with BC1 except that it has a lower affinity for oligo(dT) cellulose. BC2 is detected with variable efficiency in RNA blotting experiments because small nucleic acids do not bind quantitatively to nitrocellulose. These and further properties of BC1, BC2, and their template ID are summarized in Table 2.

We have conducted a developmental study, isolating poly(A)$^+$ RNA from the brains and also from peripheral organs of rats at days 14, 16, 18, and 20 of embryonic development and 1, 5, 10, 15, 20, 25, and 30 days after birth. BC1 and BC2 RNA's begin to be detected in hybridization experiments (Fig. 1) in the samples from around the time of birth and in large amounts beginning 5 to 10 days postnatally and are found exclusively in the brain. Our probes hybridize poorly to all prenatal brain samples, even though probes for control mRNA's hybridize well with the early brain and nonbrain samples. Work by several groups (*28*) has shown that cortical neuronal chromatin in mammals exhibits a shorter nucleosome repeat than either glial or other nonneuronal tissues. The onset of this change is early postnatal life in rats

Table 2. Summary of characteristics of ID elements and their two transcripts BC1 and BC2.

	BC1 RNA	BC2 RNA	ID DNA
Nucleotide length	160 ± 3	105 ± 5	82 + A-rich 3′ end
Abundance	~2000 transcripts per cell	~2000 transcripts per cell	10^5 copies per haploid genome
Cellular location	Cytoplasm	Cytoplasm	Within brain-specific genes in genome
Tissue of expression	Neurons	Neurons	Template in neurons for BC1 and BC2
Time of expression	~5 days postnatal and thereafter	~5 days postnatal and thereafter	
Oligo(dT) affinity	+++	+/−	

(*29*), the same time at which we see a large increase in BC1 and BC2 transcription; therefore, these events might be related. A few days thereafter, synthesis of adult brain proteins begins. For example, the brain-specific putative neurotransmitter precursor 1B236 and its mRNA both appear between postnatal days 10 and 15 in the cortex and slightly earlier in the brain stem and spinal cord (*27, 30*). We will discuss the structure of the 1B236 gene below. Some other brain-specific mRNA's we have characterized first appear in the same time frame.

We have sequenced several of the large cDNA clones that hybridize to BC1 and BC2 and found that the clones contain both common and unrelated sequences. The common sequence, which we have called an ID sequence (*21, 23*), consists of 82 nucleotides followed by an A-rich tract, and it is this element that hybridizes to BC1 and BC2 RNA's. The remaining cDNA sequences are dissimilar except that they exhibit some of the characteristics of introns or other noncoding regions, such as strings of repeated short oligonucleotides, long tracts of purines or pyrimidines, or homopolymeric tracts. Detailed mapping analyses have shown that the clones contain both intron and exon regions of brain-specific transcripts (*23*). Thus we believe that the ID-containing cDNA clones are derived by copying internal regions of mRNA precursors, which appear as a contaminant of our cytoplasmic RNA preparation. The clones probably represent cDNA copies of polymerase II transcripts.

Measurements made with the use of a phage library of rat genomic clones with inserts of, on the average, 15 kilobases show that 43 percent of the clones contain at least one ID sequence. This means that there are 0.9×10^5 to 1.5×10^5 ID copies per genome, a number in excess of the 30,000 adult brain-specific poly(A)$^+$ genes (*23*). A comparison of ten ID sequences shows that they are remarkably well conserved: the consensus in Fig. 2 shows that 51 nucleotides are invariant, 24 positions differ in only one of the ten isolates, four positions differ in two of the isolates, and the other three positions are exclusively either purine or pyrimidine (*23*). This is a much higher degree of conservation of repetitive genomic sequences than has been observed for previously described elements such as Alu sequences (*31*).

Even though the ID-containing cDNA clones appear to be copies of polymerase II transcripts, the highly conserved ID sequence is homologous to the consensus polymerase III transcription initiation sequence (Fig. 2). In fact, cDNA clones containing ID sequences function as polymerase III transcription units in vitro in HeLa S100 extracts. By cleaving the clones with a variety of restriction endonucleases, we have found that polymerase III initiation begins at position 1 of the core sequence of 82 nucleotides (*22*). This corresponds exactly to the 5′ ends of the in vivo BC1 and BC2 RNA's, as we have shown by primer extension by reverse transcriptase of ^{32}P-labeled primers derived from the 3′ end of the ID sequences (indicated by the arrow in Fig. 2). DNA sequence analysis of the extended primers gives a sequence that matches the complement of the invariant regions of the ID consensus and shows heterogeneity at the positions that are not invariant (*32*). Therefore, BC1 and BC2 are very likely to be polymerase III transcripts of genomic ID elements and are transcribed in a neuronal-specific fashion from a large number of different templates in the genome.

To resolve further the issue of whether ID sequences are transcribed in vivo by polymerase II, polymerase III, or both, we have used nuclei from brain, liver, and kidney cells to program transcription of ^{32}P-labeled RNA in a HeLa S100 system (*33*) and hybridized the transcription products to filters on which denatured ID-containing plasmid DNA's were immobilized (*22*). After correction for hybridization efficiency, we find that 62 percent of the brain transcripts, but only background amounts (<4 percent) of liver or kidney transcripts, hybridize to the filter and thus contain ID sequences. When the filter-bound RNA is eluted and examined by gel electrophoresis, two sorts of products are observed: large heterogeneous RNA and a group of small 100- to 110-nucleotide species. The synthesis of most of the large molecules is sensitive to α-amanitin (at 1 μg/ml, a concentration that inhibits polymerase II but not polymerase III transcription), whereas the synthesis of the smaller species, which are probably the BC2 species observed in vivo, is resistant to α-amanitin. Therefore, ID sequences are located in a substantial portion of brain transcripts synthesized by both polymerase II and polymerase III but are missing from liver and kidney transcripts.

Interestingly, liver and kidney program a distinct set of 100- to 110-nucleotide RNA's that do not hybridize with the ID sequence and whose synthesis is resistant to α-amanitin, suggesting analogous polymerase III transcription elements in those tissues. All of these studies, which are summarized in Table 3, give qualitatively similar results whether the transcription is conducted in HeLa S100 extracts (*33*), HeLa whole cell extracts (*34*), or with no added protein (endogenous synthesis).

If liver chromatin is treated with 0.35*M* NaCl, conditions under which nucleosomes are known to slide, transcription of RNA molecules containing the ID sequences occurs, but only the large heterogeneous species are seen. This is not true for salt-washed brain chromatin, where no qualitative change in the transcription pattern is observed although the amount of synthesis is considerably decreased. We interpret these results to mean that nucleosomes are repressing polymerase II transcription of RNA containing ID sequences in the liver, and salt wash relieves that repression. In addition, these experiments suggest that transcription by polymerase III of the small BC2 RNA requires brain-specific transcription factors (*22*). Further support for this notion comes from the observation that deproteinized genomic DNA supports the transcription of large heterogeneous molecules containing ID sequences but not transcription of BC2.

We believe that BC2 is the primary

Table 3. Nuclear transcription of ID sequences in large polymerase II (hnRNA) and small, discrete polymerase III (BC2) transcripts. Transcription reactions were carried out in the present or absence of the polymerase II inhibitor α-amanitin [see text and (*22*)], and the chromatin preparations were either native or salt-washed. Transcripts containing ID sequences (ID^+) were purified by hybridization (nonhybridizing material is indicated by ID^-) and analyzed by gel electrophoresis to determine their sizes and amounts (indicated by +, ++, +++, or −).

	Template		Product			
			hnRNA		BC2	
	α-Amanitin	NaCl wash	ID^+	ID^-	ID^+	ID^-
Brain	−	−	+++	+++	++	+++
	+	−	+	+	++	+++
	−	+	+	+	+	+
Liver	−	−	−	+++	−	+++
	−	+	+	+	−	+
Kidney	−	−	−	+++	−	++
DNA	−	−	+++	+++	−	−

polymerase III transcript of the brain ID sequence and that BC1 could be generated by occasional polyadenylation of the polymerase III transcripts. BC1 might be functional in its own right, performing some neuron-specific role, but it is not a necessary player in the formal model we suggest below. BC1 could be produced by polyadenylation of BC2 if some of the 100,000 ID sequences in the genome contain a T in their oligo(A) 3′ end, producing the AATAAA signal for cleavage and polyadenylation which is known to work on polymerase II transcripts. Other polymerase III transcripts are known to be polyadenylated (*35*).

When ID clones are used in Northern blot experiments to probe preparations of nuclear RNA made from brain, liver, and kidney, the probes hybridize much more strongly to both large $poly(A)^+$ and $poly(A)^-$ nuclear RNA from brain than to liver or kidney $poly(A)^+$ or $poly(A)^-$ nuclear RNA's (*36*). Thus, as suggested by the nuclear transcription experiments, ID sequences are found in polymerase II transcripts (heterogeneous nuclear RNA) made only in neural tissue, in the brain-specific polymerase III transcript BC2, and also in $poly(A)^-$ nuclear RNA. Hahn and his colleagues have described a class of $poly(A)^-$ RNA's in rodent brain with equivalent but nonoverlapping complexity to the class of $poly(A)^+$ RNA's. Most of these 50,000 $poly(A)^-$ RNA species appear postnatally (*37*). It is not known whether the brain-specific $poly(A)^-$ RNA's are made by polymerase II or polymerase III, but one possible explanation for their origin is that their synthesis is initiated at ID sequences by polymerase III and ends at polymerase III terminators at a distance 3′ to the ID element; hence the transcripts are not generally polyadenylated. In nuclear transcription experiments, we find the synthesis of some large RNA molecules containing ID sequences in the presence of α-amanitin. Although this may explain the origin of brain $poly(A)^-$ RNA, it does not address the possible role of these species in encoding brain-specific proteins. It does, however, offer an explanation for their postnatal appearance, since BC1 and BC2 transcription largely begins after 5 days of postnatal development, the time at which these polymerase III promoters become active.

The transcription and nuclear RNA hybridization experiments suggest that ID sequences are specifically linked to brain genes because they are found in brain but not liver or kidney primary transcripts. However, since no more than 3 percent of brain cDNA clones and brain mRNA's contain ID sequences but about 60 percent of the primary transcripts contain the ID sequence, the linkage in primary transcripts must be unstable. That is, the ID sequences are located in the introns of most adult brain-specific $poly(A)^+$ genes (at least by the mass of their transcripts), a conclusion consistent with our mapping studies of individual ID-containing cDNA clones (*23*). From the phage genomic library we isolated the gene corresponding to the 1B236 mRNA, which encodes a brain-specific precursor of a putative neurotransmitter (*27*), and used fragments from the 1B236 cDNA clone, which does not contain ID sequences, to map intron and exon regions of the gene. The 1B236 gene contains at least four introns, two of which contain ID sequences (*23*). Thus, it seems that ID sequences are highly correlated with brain-specific genes and occupy some, but not all, introns of at least most adult brain-specific genes.

There are, however, more genomic ID sequences (10^5) than $poly(A)^+$ brain-specific mRNA's (3×10^4). Some genes, such as 1B236, may contain more than one ID sequence, but other ID sequences are likely to be nonfunctional (*23*), such as those in the pseudotubulin and prolactin genes, which probably got there by transposition (*38*). We argue below that ID sequences are involved in the coordinate control of neuronal gene expression. Therefore, transposability may have provided the necessary flexibility to establish coordinate control over 30,000 genes without independently evolving control elements for each brain gene. This could help account for the rapid evolution of the mammalian brain but would be expected sometimes to generate nonfunctional or lethal copies of ID, the biological price for developing a highly sophisticated brain.

Formal Model for Lineage-Based Gene Control

The provisional model that we propose to explain these data is that ID sequences are controlling elements located within introns or adjacent to the 30,000 adult neuronal genes and act in *cis*. A factor present only in postnatal or adult neurons acts in *trans* to recognize the ID sequence and to direct its transcription by polymerase III. Either the recognition or the transcription events activate the neuronal genes in a primary manner. This gene activation is necessary but not sufficient to initiate polymerase II transcription. A second control level must exist because genes such as that encoding the 1B236 neurotransmitter protein are only active in certain brain cells: most neuronal-specific genes are probably individually expressed in small subsets of neurons. The second necessary level of gene control may operate at promoter sequences in response to ambient signals, such as cell contact and hormone or transmitter stimulation, both during postnatal development and in the adult. ID sequences, therefore, may be the neuronal lineage–specific control sequences. Although circumstantial, their correlation with neuronal-specific genes is compelling evidence that ID sequences are involved in brain-specific gene expression. The observation that ID sequences are transcribed by polymerase III in vitro from plasmid DNA and in brain nuclear extracts and that there are brain-specific ID transcription factors suggests, but does not prove, that polymerase III acting at the ID sequences might be involved in brain-specific polymerase II initiation. It could

be that the generation of a polymerase III transcript is irrelevant in the gene activation step. However, some other tissues probably have an analogous ID control system because liver and kidney nuclei generate polymerase III transcripts similar in size to BC2 but with different sequences (*22*).

An alternative version of our model, which seems more complicated to us, is that only some ID sequences might be transcribed by polymerase III in all neurons to produce the neuronal-specific BC1 and BC2 RNA's. One or both of these RNA's might then act in *trans* at the nontranscribed ID sequences to activate all neuronal genes in the primary sense. That is, brain-specific factors transcribe a few ID genes and the resulting BC transcripts activate other brain-specific genes: a sort of relay version of our basic model. Again, a second level of positive control will be necessary to initiate the transcription by polymerase II of any particular neuronal gene. This version is formally equivalent to our preferred model but differs mechanistically in that gene activation is induced by an RNA-DNA interaction rather than a protein-DNA interaction. Although we feel that this version of the model is less likely, it does provide a specific explanation for the role of polymerase III transcription in the activation of ID sequences.

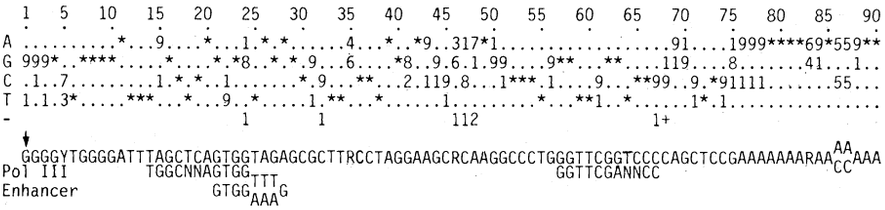

Fig. 2. The consensus ID sequence arrived at by comparing the ten sequences determined found earlier (*23, 38*). The top portion of the figure indicates, for each position, the frequency of each nucleotide [(*) 10/10, (·) 0/10, (–) gap required to align a sequence, (+) insertion required]. The common sequence is shown below, where Y indicates pyrimidine and R purine for the three positions that are not clear-cut. The arrow over the first G represents the first nucleotide of the BC1 and BC2 RNA's and also the polymerase III initiation site in vitro. The consensus polymerase III and enhancer sequences are indicated below (N indicates any base).

Mechanism of Lineage-Based Gene Control

Our formal model, which suggests that ID elements are involved in the control of gene expression, endows ID sequences with some properties similar to those described for enhancer elements. This line of logic has led us to ask whether ID sequences might be related to enhancer elements. We have substituted the ID element for the SV40 enhancer in both orientations in a plasmid series in which the neomycin resistance gene is expressed from the SV40 early promoter. In preliminary experiments, we have found that the ID sequence, operating in either orientation, is as good as the SV40 enhancer and is seven times better than enhancerless plasmids in stimulating the production of stable G418-resistant transformants of murine 3T3 and L cells (*39*). Thus ID elements can serve as enhancers. The apparent lack of cellular specificity (these are fibroblast-derived lines) is understandable because both 3T3 cells and L cells, but not other cell lines, express BC1 and BC2 RNA's and hence presumably the neuronal ID transcription factors. 3T3 cells express other neuronal-specific mRNA's as well, and hence these lines have at least some neuronal character and are not "normal" fibroblasts.

A further similarity is suggested by sequence comparison of ID and enhancer elements. The sequence $GTGG^{TTT}_{AAA}G$ or a closely related sequence is present in viral enhancers and is essential for enhancer function (*11*). Three copies of this sequence are present within the 314-bp restriction fragment of the Ig intron, which has been shown to contain the enhancer activity (*40*). The ID sequence contains a somewhat homologous sequence, GTGGTAG, in a position overlapping the 5′ or A-block region of the polymerase III split promoter (shown in Fig. 2). This could possibly be part of the site at which the neuronal-specific *trans* activator interacts with the ID sequence. Viral enhancer regions are associated with regions that can potentially form Z DNA (*16*), and junctions between B and Z DNA are preferential sites in supercoiled DNA for cleavage by S1 nuclease. The *Xenopus laevis* oocyte 5*S* gene, a developmentally regulated polymerase III transcription unit, also contains a preferential cleavage site for S1 nuclease in supercoiled plasmids located within the A block of its consensus polymerase III sequence as well as a deoxyribonuclease I cleavage site in the same region when the TFIIIA transcription factor is bound, indicating the presence of an irregular DNA configuration in this region (*41*). Further experiments are aimed at determining whether the polymerase III A block-enhancer region of ID sequences may display structural peculiarities that manifest themselves to nucleases. Such irregular DNA conformations could be relevant to ID function.

Because of the observation that adult cortical neuronal chromatin exhibits a shortened nucleosome packing density (*28, 29*), we believe there is a temporal order in which the activation of regions of chromatin by differentiation factors precedes the activation of polymerase II promoters. Implicit in this argument is that all brain ID sequences are available in all neurons. Thus a region of active chromatin is first defined by its ID sequence, and a local promoter can be recognized subsequently by the proteins required to initiate transcription. We envision that, in general, cellular genes are repressed in each tissue for polymerase II transcription by chromosome structure unless the gene contains a tissue-specific ID sequence activated by that tissue's particular factors. Unless ID activation occurs, polymerase II promoters cannot be recognized by their own *trans* activators. We must, therefore, explain how it is that ID sequences can be recognized by *trans*-acting proteins while polymerase II promoters cannot. Peculiar DNA structures of the sort discussed above may be the key. We must assume that, whatever the differentiation program is, it provides an ordered series of *trans* activators of ID type sequences. These positive regulatory factors must be able to interact with their target ID sequences, even though most DNA is unavailable for solution reactions because of its interaction with nucleosomes. It must be that the various tissue ID sequences are not inhibited by chromatin structure from protein interactions. We suggest that this is either because of peculiar DNA conformations assumed by ID-type sequences that interact differently with nucleosomes than do other regions of DNA, because ID-type sequences interact preferentially with constitutively expressed nonnucleosomal proteins that in turn exclude nucleosomes, or because differentiation factors bind to ID sequences at the time of replication and before nucleosomes form stable complexes on the gene [precedence for the latter two possibilities comes from studies on the *Xenopus* 5*S* genes (*42*)]. Either situation must leave

ID DNA available for solution reactions. We suggest that the process also might order, either actively or passively, nucleosome binding in a manner that will be operationally defined below. Thus, whenever a new stage of determination is achieved by a cell, new factors are produced that can interact with their cognate ID sequences either directly or through the specifically bound nonnucleosomal protein. We postulate that some aspect of this interaction, possibly activation of polymerase III transcription, affects local nucleosome interactions (perhaps mediated by histone acetylation), disrupting the solenoid repeating structure of chromatin (perhaps by cooperative allosteric changes of adjacent nucleosomes) and thus sending a signal over several kilobases that renders all of that DNA accessible to protein interaction. It may be that nucleosome ordering due to the events that occur at one ID sequence allows such a signal to be propagated over a domain but that the organization of nucleosomes around ID sequences in neighboring domains limits the extent of the influence of individual ID elements. Such spheres of influence might correspond to what has been referred to as active and inactive domains of chromatin. The formal features of this model are illustrated in Fig. 3.

One consequence of this model is that even genes constitutively expressed in all tissues might require ID-type sequences. These ID sequences will represent some degenerate, general form that reacts with a factor present in all cell types. The viral enhancer elements could be representative of such sequences, acting as the "lineage" markers for genes not exclusive to a specific tissue. As discussed above, they contain regions of Z DNA that may alter the way in which nucleosomes interact with the DNA and may also order their local structure. The SV40 enhancer sequence is not covered by nucleosomes, and Z DNA formation inhibits nucleosome reconstitution (*43*). If enhancers are indeed a subclass of ID sequences, since it seems unlikely that they are transcribed by polymerase III, specific protein interactions at brain ID sequences may be more important than the production of a transcript. However, it may be that important features of polymerase III promoters are the same features required for proper ID function.

A second consequence of this logic is that there will be some genes that need to be expressed in a few, but not all, tissues. These could have a small number of distinct ID-type sequences within them so that they may be active in more than one, but not all, tissues. Similarly, there may be some genes whose expression in a single tissue can be activated by more than one ambient signal. Such genes may have complex promoters. A clear advantage of our two-step gene activation model is that some ambient signals must activate different genes in different tissues. This can only be possible if the particular gene targets for the ambient signals have been independently selected in each cell type so that only certain genes can be induced in a particular cell.

Although our model for a lineage-based control system for tissue-specific genes is derived from the study of the transcripts of one tissue, the brain, it seems logical that, if the model is correct, a similar system will work for other tissues. We must stress, however, that this is only a model. Because it makes many testable predictions, we hope that it will stimulate research. We expect that most genes, tissue-specific or general, will contain *cis* regulators that determine whether they occupy "active regions" of chromatin in each particular cell type. Some of these will be polymerase III promoters, as we expect to find in liver and kidney, but others may act through different mechanisms. Enhancers may fall into this class of element. The *trans*-acting factors that govern the function of these *cis*-acting sequences must be the primary phenotypic markers of determination or lineage differentiation. The control of the expression of these factors is differentiation.

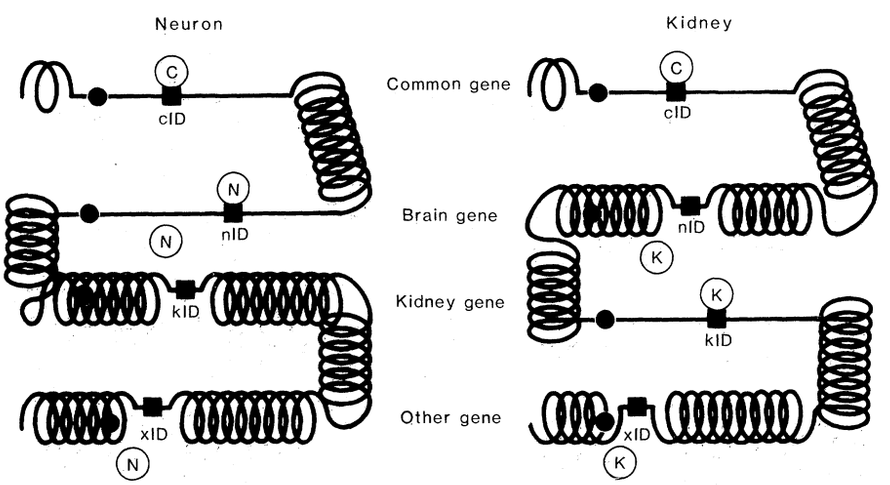

Fig. 3. Cartoon model for the formal concepts of lineage-based gene control. Four types of genes are illustrated: a gene expressed in all tissue types, a neuronal-specific gene, a kidney-specific gene, and a gene specific for a third (x) tissue. Their states are shown in two different cell types: neuronal (N) and kidney (K). Promoter regions are indicated as solid circles; tissue-specific positive *trans* activators are shown as lettered white circles. In all cases the DNA interacts with nucleosomes, except possibly at the various types of ID regions (which may be nucleosome-free), one of which is shown in each gene (c, common; n, neuronal; k, kidney; x, another tissue); however, the quality of the nucleosome-DNA interactions are different in the active and inactive genes. The solenoidal structures represent higher order supernucleosomal levels of structure, while the straight lines represent extended chromatin structures typical of active genes. Because transcription of a gene in the "active" lineage state is conditional on the proper polymerase II factors being present, no transcripts are indicated in the figure. Similarly, because we do not know whether polymerase III transcriptional events are crucial to neuronal ID function, ID transcripts are not shown.

References and Notes

1. R. J. Britten and E. H. Davidson, *Science* **165**, 349 (1969); *Quart. Rev. Biol.* **46**, 111 (1969).
2. H. R. B. Pelham, *Cell* **30**, 517 (1982); ——— and M. Bienz, *EMBO J.* **1**, 1473 (1982).
3. C. Parker and J. Topol, *Cell* **37**, 273 (1984).
4. A. Schurch, J. Miozzari, R. Hutter, *J. Bacteriol.* **117**, 1131 (1974); J. Delforge, F. Messenguy, J. Wiame, *Eur. J. Biochem.* **57**, 231 (1975).
5. T. F. Donahue, R. S. Daves, G. Lucchini, G. R. Fink, *Cell* **32**, 89 (1983); A. G. Hinnesbusch and G. R. Fink, *J. Biol. Chem.* **258**, 5238 (1983); A. G. Hinnebusch, G. Lucchini, G. R. Fink, in *Enhancers and Eukaryotic Gene Expression*, Y. Gluzman and T. Shenk, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1983), p. 194.
6. W. S. Dynan and R. Tijan, *Cell* **32**, 669 (1983); *ibid.* **35**, 79 (1983).
7. F. Payvar *et al.*, *J. Cell Biochem.* **19**, 241 (1982); V. L. Chandler, B. A. Maler, K. R. Yamamoto, *Cell* **33**, 489 (1983); F. D. Payvar *et al.*, *ibid.* **35**, 381 (1983).
8. J. Banerji, S. Rusconi, W. Schaffner, *Cell* **27**, 299 (1981); P. Gruss, R. Dhar, G. Khoury, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 943 (1981); P. Moreau, R. Hen, B. Wasylyk, R. Everett, M. P. Gaub, P. Chambon, *Nucl. Acids Res.* **9**, 6047 (1981); R. K. Humphries, T. Ley, P. Turner, A. D. Moulton, A. W. Nienhuis, *Cell* **30**, 173 (1982).
9. C. Benoist and P. Chambon, *Nature (London)* **290**, 304 (1981).
10. M. Fromm and P. Berg, *J. Mol. Appl. Genet.* **1**, 457 (1982).
11. H. Weiher, M. Konig, P. Gruss, *Science* **219**, 626 (1983).
12. J. de Villiers and W. Schaffner, *Nucl. Acids Res.* **9**, 6251 (1981).
13. M. Lusky, L. Berg, H. Weiher, M. Botchan, *Mol. Cell. Biol.* **3**, 1108 (1983).
14. L. A. Laimins, G. Khoury, C. Gorman, B. Howard, P. Gruss, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 6453 (1982).
15. S. Gillies, S. Morrison, V. Oi, S. Tonegawa, *Cell* **33**, 718 (1983).
16. A. Nordheim and A. Rich, in *Enhancers and Eukaryotic Gene Expression*, Y. Gluzman and

T. Shenk, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1983), p. 45.
17. G. Levinson, G. Khoury, G. Vande Woude, P. Gruss, *Nature (London)* **295**, 568 (1982).
18. H. Scholer and P. Gruss, in *Enhancers and Eukaryotic Gene Expression*, Y. Gluzman and T. Shenk, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1983), p. 16.
19. S. Gillies and S. Tonegawa, *Nucl. Acids Res.* **11**, 7981 (1983); S. Gillies, S. Morrison, V. Oi, S. Tonegawa, in *Enhancers and Eukaryotic Gene Expression*, Y. Gluzman and T. Shenk, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1983), p. 121; J. Baneji, L. Olson, W. Schaffner, *ibid.*, p. 136.
20. A. R. Kimmel, C. Lai, R. Firtel, in *ICN-UCLA Symposium on Molecular and Cellular Biology*, R. Axel, T. Maniatis, C. F. Fox, Eds. (Academic Press, New York, 1979), vol. 14, p. 195; A. R. Kimmel and R. Firtel, *Cell* **16**, 787 (1979).
21. J. G. Sutcliffe, R. J. Milner, F. E. Bloom, R. A. Lerner, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 4942 (1982).
22. J. G. Sutcliffe, R. J. Milner, J. M. Gottesfeld, R. A. Lerner, *Nature (London)* **308**, 237 (1984).
23. R. J. Milner, F. E. Bloom, C. Lai, R. A. Lerner, J. G. Sutcliffe, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 713 (1984).
24. P. Brickell, D. Latchman, D. Murphy, K. Willison, P. Rigby, *Nature (London)* **306**, 756 (1983).
25. W. Klein, T. Thomas, C. Lai, R. Scheller, R. Britten, E. Davidson, *Cell* **14**, 889 (1978); J. W. Posakony, G. Flytzanis, R. Britten, E. Davidson, *J. Mol. Biol.* **167**, 361 (1983).
26. R. J. Milner and J. G. Sutcliffe, *Nucl. Acids Res.* **11**, 5497 (1983).
27. J. G. Sutcliffe, R. J. Milner, T. M. Shinnick, F. E. Bloom, *Cell* **33**, 671 (1983); J. G. Sutcliffe and R. J. Milner, *Trends Biochem. Sci.* **9**, 95 (1984).
28. J. O. Thomas and R. J. Thompson, *Cell* **10**, 633 (1977).
29. P. D. Greenwood and I. R. Brown, *Neurochem. Res.* **7**, 965 (1982); I. R. Brown, *Biochim. Biophys. Acta* **698**, 307 (1982); C. C. Kuenzle *et al.*, *Cold Spring Harbor Symp. Quant. Biol.* **47**, 493 (1983).
30. D. Lenoir, E. L. F. Battenberg, F. E. Bloom, P. Danielson, J. G. Sutcliffe, R. J. Milner, unpublished observations.
31. P. L. Deininger, D. J. Jolly, C. M. Rubin, T. Friedmann, C. W. Schmid, *J. Mol. Biol.* **151**, 17 (1981).
32. J. Arriza, A.-P. Chen, P. Danielson, C. Lai, R. J. Milner, J. G. Sutcliffe, unpublished observations.
33. P. A. Weil, J. Segall, B. Harris, S.-Y. Ng, R. G. Roeder, *J. Biol. Chem.* **254**, 6163 (1979).
34. J. L. Manley, A. Fire, A. Cano, P. A. Sharp, M. L. Gefter, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 3855 (1980).
35. D. P. Carlson and J. Ross, *Cell* **34**, 357 (1983).
36. W. Hahn, personal communication.
37. N. Chaudhari and W. E. Hahn, *Science* **220**, 924 (1983); W. E. Hahn, N. Chaudhari, L. Beck, K. Wilber, D. Peffley, *Cold Spring Harbor Symp. Quant. Biol.* **48**, 465 (1983).
38. I. Lemischka and P. Sharp, *Nature (London)* **300**, 330 (1982); L. Schuler, J. Weber, J. Gorski, *ibid.* **305**, 149 (1983); F. G. Gonzalez and C. B. Kasper, *J. Biol. Chem.* **258**, 1363 (1983).
39. R. McKinnon, T. M. Shinnick, J. G. Sutcliffe, unpublished observations.
40. J. Banerji, L. Olson, and W. Schaffner, in *Enhancers and Eukaryotic Gene Expression*, Y. Gluzman and T. Shenk, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1983), p. 136.
41. W. Reynolds and J. Gottesfeld, in preparation.
42. J. Gottesfeld and L. S. Bloomer, *Cell* **28**, 781 (1982); D. F. Bogenhagen *et al.*, *ibid.*, p. 413.
43. S. Saragosti, G. Moyne, M. Yaniv, *ibid.* **20**, 65 (1980); E. B. Jakobovitz, S. Bratosin, Y. Aloni, *Nature (London)* **285**, 263 (1980); A. J. Varshavsky, O. H. Sundin, M. J. Bohn, *Cell* **16**, 453 (1979); J. Nickol, M. Behe, G. Felsenfeld, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 1771 (1982).
44. We thank R. McKinnon, T. Shinnick, D. Pallotta, F. Bloom, R. Lerner, J. Arriza, A.-P. Chen, P. Danielson, and C. Lai for numerous discussions and collaborations on unpublished experiments; B. Hahn for communicating unpublished results; and L. Eddy for manuscript assistance. Supported in part by a grant from the National Institutes of Health (GM32355) and McNeil Laboratories. Publication 3529-MB from the Research Institute of Scripps Clinic.

Alternative RNA Processing: Determining Neuronal Phenotype

Michael G. Rosenfeld, Susan G. Amara, Ronald M. Evans

Summary. On the basis of an analysis of the human and rat calcitonin genes and of a related gene, alternative RNA processing represents a developmental strategy of the brain to dictate tissue-specific patterns of polypeptide synthesis. This regulation allows the calcitonin gene to generate two messenger RNA's, one encoding the precursor of a novel neuropeptide, referred to as CGRP, which predominates in the brain, and the second encoding the precursor to the hormone calcitonin which predominates in thyroid C cells. The distribution of CGRP in the central and peripheral nervous system and in endocrine and other organ systems suggests potential functions in nociception, ingestive behavior, cardiovascular homeostasis, and mineral metabolism.

Complex regulatory mechanisms restrict the expression in neural tissues of genes encoding neuroendocrine peptides to precise groups of neurons. Understanding the developmental mechanisms responsible for such regulation is likely to provide general insights into the molecular strategies critical for brain development and function. Regulation of gene expression in the brain, as in other tissues, could occur during gene transcription, RNA processing and transport, messenger RNA (mRNA) stability, and posttranslational events (*1*). Because RNA processing regulation has the potential to qualitatively, as well as quantitatively, alter the nature of the gene product, it would present specific advantages in expression of certain genes. Analysis of calcitonin gene expression in neural tissues suggests that differential RNA processing events are one type of developmental regulation specifying the pattern of neuroendocrine gene expression (*2–6*), and could represent a common strategy in expression of certain genes dictating neural development. Alternative RNA processing occurs in a tissue-specific fashion to produce alternative polypeptide products (*5, 6*) and, therefore, serves to increase the diversity of neuropeptides generated from a single genomic locus.

A Single Neuroendocrine Gene Generates Multiple RNA Products

The preparation and DNA sequence analysis of plasmids containing DNA complementary to rat calcitonin mRNA predicted the structure of the protein precursor to calcitonin, a 32–amino acid calcium-regulating hormone, produced in thyroid C cells (Fig. 1) (*2, 7, 8*). In addition to calcitonin, proteolytic processing of the precursor generates in thyroid C cells an 82–amino acid NH_2-terminal peptide and a 16–amino acid COOH-terminal calcitonin cleavage product (CCP) (Fig. 1). The production of multiple calcitonin-related mRNA's was first noted during the spontaneous and permanent "switching" of serially transplanted rat medullary thyroid carcinomas (MTC's) from states of high to low or absent calcitonin production (*3, 9*). The unexpected explanation for the "switch" was that calcitonin gene transcription continued but generated a series of new, structurally distinct tran-

M. G. Rosenfeld is a professor and S. G. Amara is a Schering-Plough fellow of the Life Science Research Foundation in the Eukaryotic Regulatory Biology Program, University of California San Diego School of Medicine, La Jolla 92093. R. M. Evans is a senior member of the Molecular Biology and Tumor Virology Laboratory, Salk Institute, San Diego, California 92138.