Order Number 9304701

BC1 and BC200 RNA: Sequence and transcriptional analysis

Martignetti, John Attilio, Ph.D.

City University of New York, 1992

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

BC1 AND BC200 RNA:

SEQUENCE AND TRANSCRIPTIONAL ANALYSIS

by

John A. Martignetti

A dissertation submitted to the Graduate Faculty in Biomedical Sciences in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

1992

This manuscript has been read and accepted for the Graduate Faculty in Biomedical Sciences in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

8/17/92
Date

Chairman of Examining Committee

8/19/92
Date

Supervisor

8/19/92
Date

Executive Officer

Supervisory Committee: James Bieker
Nouria Hernandez
Stave Kohtz
Stephen Salton

THE CITY UNIVERSITY OF NEW YORK

PREFACE

This thesis has not been submitted in whole or in part for a degree or other qualification at any other university. This thesis is substantially my own work. Where reference has been made to other research this is acknowledged in the text and bibliography.

To my parents, yet once again, thank you.

To Ilene Rabinowitz, thank you for your loving support.

To Jürgen Brosius, thank you for your patience, advice, sincere concern, and overall guidance. To members, past and present, of the Brosius Lab, K. Eisinger, B. Probst, M. Miller, H. Tiedge, Jr-Gang Cheng, W. Chen, and S. Kaplan, thank you for making these years even more enjoyable and productive.

To Jim Bieker, Nancy Levin, Shona Murphy, and Steve Salton, thank you all for help with problem solving and of course, "hand-holding", during periods of this research, where without your guidance and support this work would have faltered. To Josh Berman, Shari Dermer, Tina Lopingco, and Manami Tsutsumi, thank you for your friendship without which these years would not have been as enjoyable.

To James Roberts and T. Krulwich, thank you for your faith in me and all the problems it must have caused.

To Marc Glucksman, of course, thank you for all of the above, but most of all thank you for being such a good friend.

Nel mezzo del cammin di nostra vita
mi ritrovai per una selva oscura
che la diritta via era smarrita.
Ah quanto a dir qual era é cosa dura
esta selva selvaggia e aspra e forte
che nel pensier rinova la paura!
Tant' é amara che poco é piu morte;
ma per trattar del ben ch'io vi trovai,
diro dell'altre cose ch'i' v'ho scorte.

In the middle of the journey of our life I came to myself within a dark wood where the straight way was lost. Ah, how hard a thing it is to tell of that wood, savage and harsh and dense, the thought of which renews my fear! So bitter is it that death is hardly more. But to give account of the good which I found there I will tell of the other things I noted there.

(The Divine Comedy, Dante)

TABLE OF CONTENTS

THESIS ABSTRACT....1

A NOTE TO THE READER....3

CHAPTER 1 INTRODUCTION

Identification and structure of BC1 and BC200 RNAs....4
Tissue distribution and expression....8
An originally defined role for BC1 RNA....11
A possible function for BC1 and BC200 RNAs....14
Interspersed repetitive DNA....15
Transcriptional regulation....18
RNA polymerase I....18
RNA polymerase II....19
RNA polymerase III....20
Transgenic mice....23
DNA sequencing....25

CHAPTER 2 MATERIALS AND METHODS

Materials....29
Methods....32
DNA sequencing and analysis....32
Template preparation....32
Sequencing reactions....34
Data analysis....35
Guinea pig Southern and Northern blot analyses....35
PCR labeling of oligonucleotides....36

CHAPTER 3 RESULTS

Organization, structure and sequence analysis of the BC1 RNA gene of rodents....38
The neural specific expression of BC1 RNA has been conserved for 55 million years....50

The establishment of rat BC1 RNA transgenic mice lines.....55

CHAPTER 4 EVOLUTIONARY IMPLICATIONS OF THE CONSERVATION OF THE BC1 RNA GENE.....63

CHAPTER 5 THE NEURAL SPECIFIC BC1 RNA GENE: AN RNA POLYMERASE III TRANSCRIPT WITH FUNCTIONAL POL II AND POL III PROMOTER ELEMENTS.....79

CHAPTER 6 A NEURAL SPECIFIC RNA POLYMERASE III PRODUCT ENCODED BY A MONOMERIC ALU ELEMENT.....135

CONCLUSIONS.....151

REFERENCES.....157

APPENDIX.....168

-BC1 RNA GENES

Rat.....A
Mouse.....C
Chinese Hamster.....D
Guinea Pig.....E
Syrian Hamster Pseudogene.....F

-HUMAN BC200 RNA GENES

BC200α.....G
BC200β.....J
BC200γ.....N

-BC200β RNA PSEUDOGENE LOCUS

Chimpanzee.....O

LIST OF FIGURES

CHAPTER 1 INTRODUCTION

The structure of BC1 and BC200 RNAs....6

CHAPTER 3 RESULTS

Diagram of the rodent BC1 RNA genes and a BC1 RNA pseudogene....39

Pairwise comparisons of the rodent BC1 RNA gene and flanking sequences....42

Alignment of the BC1 RNA coding region of several rodent species....44

The 5' flanking regions of the BC1 RNA gene of three rodent species....46

The 3' flanking regions of the BC1 RNA gene of three rodent species....48

Southern analysis of the guinea pig BC1 RNA gene....51

Guinea pig BC1 RNA is transcribed in a neural specific manner....53

Chimaeric constructs used for transgenic studies....56

Agarose gel electorphoresis of transgenic tail samples....59

Southern analysis of transgenic DNA samples....61

CHAPTER 4 EVOLUTIONARY IMPLICATIONS OF THE CONSERVATION OF THE BC1 RNA GENE

Northern blots of rat, rabbit, bovine and human brain and various guinea pig RNAs....73

cDNA sequence alignment of BC1 RNA from rat, mouse, syrian golden hamster and guinea pig....75

Phylogenetic trees for guinea pig, myomorph rodents, artiodactyls, primates and an outgroup....77

CHAPTER 5 THE NEURAL SPECIFIC BC1 RNA GENE: AN RNA POLYMERASE III TRANSCRIPT WITH

FUNCTIONAL POL II AND POL III PROMOTER ELEMENTS

Effect of $MgCl_2$ and temperature on the *in vitro* transcription reaction....117

The sequence of the BC1 RNA upstream and coding region....119

The BC1 intragenic promoter sequences are required for *in vitro* transcription....121

The BC1 b' box and a consensus B box are not mutually interconvertable....123

The effect of interaction between intragenic and extragenic sequences on *in vitro* synthesis....125

Mapping the BC1 RNA gene upstream sequences that regulate *in vitro* transcription....127

The BC1 TATA and PSE elements are functional in regulating BC1 RNA synthesis....129

Schematic model for the interaction between intragenic and extragenic factors in pol III transcription....131

CHAPTER 6 A NEURAL SPECIFIC RNA POLYMERASE III PRODUCT ENCODED BY A MONOMERIC ALU ELEMENT

Structure of the human BC200α, β and γ genes....145

Sequence alignment of the BC200 RNA coding region....147

In vitro transcription of the BC200α gene....149

THESIS ABSTRACT

Rodent BC1 RNA and primate BC200 RNA are tissue specific, nontranslatable small cytoplasmic RNAs which represent the first examples of neural specific RNA polymerase III transcribed products. BC1 RNA evolved via a $tRNA^{Ala}$ and BC200 RNA via 7SL RNA. Despite their seemingly unrelated phylogenetic origins, these two molecules share a similar tripartite structure and limited sequence identity in their 3' ends. Furthermore, these RNAs, in the form of ribonucleoprotein complexes (RNP), are present not only in the somatic but also the dendritic compartments of analogous subsets of neurons. These results, coupled with BC1 RNA's expression coinciding with developmental synaptogenesis and BC200 RNA's ability to bind two protein subunits of the signal recognition particle, suggest that BC1 and BC200 RNAs are involved in postsynaptic protein biosynthesis.

The genes encoding BC1 RNA from several rodent species, including rat, mouse, Chinese hamster and guinea pig, and human BC200 RNA have been identified, sequenced and analyzed. Additionally a number of BC1 and BC200 RNA pseudogenes have also been characterized. Analysis of these sequences revealed that although BC1 RNA is of a relatively recent phylogenetic origin, evolutionary pressure has conserved the genes and their neural specific expression for at least 55 million years. Also demonstrated is that the human BC200 RNA gene and its RNA product, respectively, represent the only known transcriptionally active Alu element and the first example of a primate RNA polymerase III transcribed neural specific RNA. Furthermore,

the BC200 RNA gene provides the first evidence of transcriptionally active monomeric Alu elements: possible progenitors to the nearly half million Alu sequences found dispersed throughout the human genome.

The second aspect of this work was the analysis of the transcriptional regulation of the BC1 RNA gene. For *in vivo* studies, several transgenic mouse lines were established. For *in vitro* studies, a novel homologous whole cell extract, prepared from rat brain tissue, capable of supporting RNA polymerase III (pol III) catalyzed transcription was developed. Using this extract, functional RNA pol II and pol III promoter elements involved in regulating rat BC1 RNA transcription were defined. Moreover, a variant pol III promoter that is able to "communicate" with pol II promoters has been identified. This suggests a mechanism by which the BC1 RNA gene and possibly other pol III transcribed genes can interact with pol II factors and be differentially regulated.

A NOTE TO THE READER

The structure of this work is such that three of the chapters presented are either currently submitted manuscripts or are manuscripts in preparation. These include Chapters 4, 5 and 6. Owing to the Materials and Methods and Reference sections contained within each, these will not be reiterated in the body of the thesis unless specifically stated.

CHAPTER 1

INTRODUCTION

> Nothing can ever be created by divine power out of nothing. ...First then, if things were made out of nothing, any species could spring from any source and nothing would require seed. Men could arise from the sea and scaly fish from the earth, and birds could be hatched out of the sky. ...The same fruits would not grow constantly on the same trees, but they would keep changing: any tree might bear any fruit. ...since each is formed out of specific seeds, it is born and emerges into the sunlit world only from a place where there exists the right material... because each thing requires for its birth a particular material which determines what can be produced.
>
> (On the Nature of the Universe, Lucretius)

Identification and structure of BC1 and BC200 RNAs

Rat BC1 RNA is a small cytoplasmic RNA, selectively and differentially expressed in brain and transcribed by RNA polymerase III (Sutcliffe et al., 1982 and 1984 a, b; Milner et al., 1984). It was originally identified on Northern blots by a probe containing a middle repetitive

sequence, the ID element. This element, represented in the rat genome 10^5 times, consists of a 75 nucleotide (nt) sequence terminating in a poly(A) tail of variable length. BC1 RNA, as revealed by Northern analysis, was shown to be a 160 nt long transcript. The difference in length between BC1 RNA and ID elements was believed attributable to BC1 RNA representing a transcriptional by-product of sequences containing these repetitive elements.

DeChiara and Brosius (1987), in an attempt to determine if BC1 RNA represents transcriptional by-products or possibly the independent transcript of a BC1 RNA gene(s), compared the sequences of 10 cDNA clones derived from *in vitro* C-tailed RNA. The rationale to utilize C-tailed RNA as a primer for reverse transcription was that it was undetermined whether the A-rich region would represent the middle or the 3' end of the BC1 RNA. If the A-rich region were in the middle of the molecule and the reverse transcription utilized oligo d(T), a full-length cDNA copy would not be generated. The clones revealed BC1 RNA to possess a tripartite structure: a 5' ID homologous portion, a central A-rich region and a terminal 25 nt section, termed BC1 unique (Figure 1A). This 25 nt unique sequence, when used as a probe in Northern blot experiments, detects only BC1 RNA and when used for Southern blots of rat genomic DNA detects only a single band. That there was both a lack of sequence heterogeneity in the cDNA clones and that only one band was detected on genomic Southerns suggests that BC1 RNA is the product of a single gene.

Watson and Sutcliffe (1987) attempted to define an homologous RNA in primates. They screened RNA derived from monkey brain, under very low stringency conditions, using the rat ID element

Figure 1. The structure of BC1 and BC200 RNAs. **A**. The primary structure of rat BC1 RNA. The sequence has been separated to detail its tripartite structure. The RNA polymerase III internal promoter sequences (boxes A and B) have been underlined and the derived consensus sequence for box A and B promoters is shown. Note that the BC1 box B sequence differs from the canonical sequence by lacking an invariant A residue. **B**. The primary structure of human BC200 RNA. As in **A**, the sequence has been separated to reveal the tripartite structure and the putative internal promoter sequences have been underlined.

A.

GGGGTTGGGGATTTAGCTCAGTGGTAGAGCGCTTGCCTAGCAAGCGCAAGGCCCTGGGTTCGGTCCTCAGCTCCG — ID region
(TGGCNNAGTGG) (GGTTCGANNCC)

AAAAAAAAAAAAAAAAAAAAAAGACAAAATAACAAAAAGACCAAAAAAAAA — A-rich region

CAAGGTAACTGGCACACACAACCTTT — unique region

B.

GGCCGGGCGCGGTGGCTCACGCCTGTAATCCCAGCTCTCAGGGAGGCTAAGAGGCGGGAGGATAGCTTGAGCCCAGGAGTTCGAGACCTGCCTGGGCAAT

ATAGCGAGACCCCGTTCTCCAG — Alu region

AAAAAGGAAAAAAAAAAACAAAAGACAAAAAAAAAA — A-rich region

TAAGCGTAACTTCCCTCAAAGCAACAACCCCCCCCCCCCTTT — unique region

containing probe which had originally detected BC1 RNA. A 200 nt RNA expressed in brain but not in nonneural tissues was identified. Partial cDNA clones revealed that this RNA shares a high degree of homology, not to the ID element, but to the left half of the human Alu element. The authors suggested that BC200 RNA may result from posttranscriptional processing of Alu containing heterogeneous nuclear RNA or that BC200 RNA may be the product of only a few genes.

That BC200 RNA may be the product of a single gene was suggested by the work of Tiedge et al., (1992). Using the same techniques as used for the isolation of BC1 RNA (DeChiara and Brosius, 1987), a number of overlapping human BC200 RNA cDNA clones were isolated and sequenced. Similarly to BC1 RNA, BC200 can be subdivided into three structural domains: these include a 5' portion homologous to the Alu left monomer, a central A-rich region, and a terminal 43 nucleotide section, named BC200 unique (Figure 1B). As with BC1 RNA, the unique 3' sequence when used as a probe on Northern blots detects only BC200 RNA. When used for Southern blots of human genomic DNA, two major hybridizing bands are detected; one the BC200 gene, the other a BC200 RNA pseudogene (work presented herein).

Tissue distribution and expression

The distribution of BC1 RNA in the adult rat central and peripheral nervous system has been revealed by *in situ* hybridization techniques (Tiedge et al., 1991). The labeling pattern revealed a distribution seemingly unrelated to specific cell types or

transmitter/receptor systems; instead, high levels of expression are restricted to gray matter areas. These areas include the neocortex, several thalamic and hypothalamic nuclei, the amygdala, the ventral lateral geniculate nucleus and superior colliculus, several other brainstem areas and the spinal cord. Thalamic nuclei strongly labelled include the paraventricular and paratenial thalamic nuclei and the medial habenular nucleus of the epithalamus. Similarly labelled hypothalamic nuclei included the supraoptic nucleus, the paraventricular hypothalamic nucleus, the dorso- and ventralmedial hypothalamic nuclei and several preoptic nuclei. In the amygdaloid complex, nuclei of the olfactory, medial, central, basolateral amygdala, the corpus striatum (moderate labeling), the bed nucleus of the stria terminalis, and the septal nuclei were all highly labelled.

Only low to moderate amounts of BC1 RNA are present in the cerebellum. White matter areas of the brain, including the lateral olfactory tract, the optic nerve, the anterior and posterior commisure, the internal capsule, the sensory root of the trigeminal nerve, and the pyramidal tract, also revealed no BC1 RNA. The absence of BC1 RNA from these regions indicates the absence of BC1 RNA from glial cells and the axons of neurons. Similarly, BC1 RNA is absent from non-neural tissues, including liver, lung, kidney, spleen and both skeletal and cardiac muscle.

Analysis of the distribution of BC1 RNA was also made at the cellular level (Tiedge et al., 1991). *In situ* hybidization experiments in the hippocampus, neocortex and the olfactory bulb as well as in acutely isolated spinal cord neurons demonstrated that BC1 RNA is present not only in the somatic but also dendritic fields of a subset of neurons. This

unusual dendritic subcellular localization is only known to be shared with two messenger RNAs, MAP2 (Garner et al., 1988; Kleiman et al., 1991; Bruckenstein et al., 1991) and the α subunit of Ca^{2+}/calmodulin dependent protein kinase type II (Burgin et al., 1990).

Rat BC1 RNA is already expressed at embryonic day 14 (E14) in the developing spinal cord. At E15/16 BC1 RNA is detectable in several brainstem areas. In the olfactory bulb, BC1 RNA levels rise significantly at E17/18 and reach peak levels shortly after birth. In the neocortex, significant BC1 RNA levels are detectable at time of birth, but during the ensuing weeks significant increases are seen, with constant levels reached in the mature brain. In the hippocampus, similar expression patterns are seen, and BC1 RNA levels begin to rise at the end of the first postnatal week. In the retina, significant BC1 RNA levels are observed at postnatal day 12 (P12) and mature levels by P19. In the cerebellum, where BC1 RNA is expressed appreciably only in the Purkinje cells, maximum levels are reached at the end of the first postnatal week. Expression then declines to the relatively low levels found in the mature animal. Interestingly, the onset of BC1 RNA expression in all of these areas coincides with the time course of synaptogenesis.

The distribution of human BC200 RNA in selected neural tissues has also been examined by *in situ* hybridization (Tiedge et al., 1992). In the tissues examined, labeling is analogous to BC1 RNA expressing. For example, high levels were observed in the neocortex, whereas no labeling was observed in white matter areas. Moreover, human BC200 RNA was localized to dendrite-rich neuropil areas as was BC1 RNA. The onset of expression of BC200 RNA during fetal and postnatal development has not yet been examined. Therefore whether the onset of

BC200 RNA expression will be correlated, as is BC1 RNA's expression, with developmental synaptogenesis is undetermined.

The originally defined role for BC1 RNA

When originally identified, BC1 RNA was believed to represent a heterogeneous population of molecules derived from the ID elements (Sutcliffe et al., 1982 and 1984 a, b; Milner et al., 1984). Owing to BC1 RNA's neural specificity and the presence of ID elements specifically within the introns of neural genes, a theory that ID elements directed brain specific gene expression arose. The ID theory presented a variation on the methods proposed by Britten and Davidson (1969; 1979) by which repetitive elements could coordinate transcriptional regulation. Much attention was focused on this mechanism for gene regulation and unfortunately the unique nature of BC1 RNA was overlooked. The ID model is therefore presented to summarize the data upon which it was founded and to clarify the distinction between ID elements and BC1 RNA.

The brain is estimated to express approximately 5×10^4 different mRNA's (Bantle and Hahn, 1976; Hastie and Bishop, 1976; Chikaraishi, 1979); greater than any other tissue. To initiate studies directed at understanding how this level of complexity is controlled, Milner and Sutcliffe (1983) created a cDNA library from rat cytoplasmic brain poly $(A)^+$ RNA to characterize clones which hybridized specifically to brain but not kidney or liver RNA. Five of the 163 neural-specific clones, all greater than 500 nts in length, hybridized to a 160 nt long RNA species

(Sutcliffe *et al.*, 1982). Additional Northern blots revealed that this 160 nt RNA possessed a variable distribution within the adult rat brain: regions analyzed included hippocampus, caudate, hypothalamus, cortex, olfactory bulb, cerebellum and pons/hindbrain.

Two of the five cDNA clones were sequenced to determine why the 160 nt RNA was detected. These two clones shared a 75 nt non-coding sequence near their 3' ends followed by a poly(A) tail. Sequences 5' to the shared element were unrelated. The 75 nt sequence was shown to be responsible for hybridization to the brain specific 160 nt cytoplasmic RNA. Analysis of the sequence database tantalizingly revealed an exact match of the 75 nt sequence to intron I_B of the rat growth hormone gene. The authors noted that: 1. the 75 nt sequence was organized similarly to Alu-like sequences in that they both terminated with variable lengths of poly(A) and hybridized to a small RNA; and, 2. the 75 nt sequence was comparable to the 72 bp enhancer elements found in SV40 that had recently been characterized. The authors termed the 75 nt sequence an "identifier (ID) sequence" and suggested that the ID sequence, found specifically in the introns and/or non-coding regions of neural genes, somehow promotes neural specific gene expression.

Milner *et al.*, (1984) showed that the ID element was: highly conserved amongst different clones and sources, contained in the introns of brain specific genes, represented in the rat genome 10^5 times, and contained sequences that were similar to pol III box A and B promoters. However, while showing that introns of brain-specific genes contain ID sequences, these authors never demonstrated that the introns of non brain specific genes do not contain ID sequences.

Sutcliffe *et al.*, (1984b) wrongly concluded from an *in vitro* system that the ID element is present in 62% of brain transcripts but virtually absent from kidney and liver transcripts. This result, coupled with their finding that the ID's internal promoter sequences directed *in vitro* pol III transcription led the authors to infer that the pol III transcription of the neural-specific ID sequences regulated the pol II tissue-specific transcription of these genes. In the model proposed, neural specific genes contained ID elements within their introns. These sequences were then recognized and transcribed by pol III. This resulted in the brain specific accumulation of the 160 nucleotide poly $(A)^+$ RNA molecule, named BC1 RNA (to discriminate it from the ID elements and to emphasize its specific localization in brain cytoplasm). The recognition of the ID DNA sequence by pol III or an interaction with BC1 RNA somehow activated RNA polymerase II to initiate transcription of the brain specific gene containing the ID sequence. In this way, possession of an ID sequence in the intronic region of a gene coordinated tissue specific transcriptional regulation.

The ID model was based on two observations: 1. the ID sequence is found solely in brain-specific pol II genes; and, 2. BC1 RNA is specifically expressed in the brain. The first observation, that the ID sequence is specific to neuronal messages, has been shown to be false. Non-neural tissues transcribe approximately equal amounts of ID containing sequences (Witney and Furano, 1984; Owens et al., 1985; Lone et al., 1986). The second observation that BC1 RNA *in vivo* is found in nervous tissue has however been substantiated (Sutcliffe et al., 1984; Owens *et al.*, 1985; Sapienza and St-Jacques, 1986; Lone *et al.*, 1986; DeChiara and Brosius, 1987). Therefore, the role of ID elements (and

BC1 RNA) in directing neural specific expression no longer seems plausible.

A possible function for BC1 and BC200 RNAs

Long-term structural and functional modulation of synaptic connections is believed to require the *de novo* synthesis of proteins (Davis and Squire, 1984; Montarolo et al., 1986). Local postsynaptic protein biosynthesis (reviewed in Steward et al., 1988; Banker and Steward, 1992) would provide a method for efficient and flexible synaptic modulation. Strongly supporting this theory are experiments which demonstrate the ability of dendrites isolated from cultured hippocampal neurons or synaptosomes to support active protein biosynthesis (Torre and Steward, 1991; Rao and Steward, 1991). Further experimental observations are also in accord with this theory: 1. polyribosomes have been shown to exist beneath synaptic sites, most prominently at the base of the dendritic spines (Steward and Levy, 1982, Steward and Reeves, 1988); 2. dendritic polyribosomes are known to be increased during periods of developmental and reactive synaptogenesis (Steward, 1983; Steward and Falk, 1986); and, 3. RNA is actively transported in the dendrites but not axons of hippocampal cells in culture (Davis et al., 1987; Davis et al., 1990).

Postsynaptic protein biosynthesis requires that components of the translational apparatus be localized within the same subcellular compartments: synapses being present in both the dendrites and also soma of neurons. Rodent BC1 and primate BC200 RNAs and their

associated RNPs possess the necessary highly unusual dendritic and somatic distribution. Furthermore, the ancestry of both BC1 and BC200 RNAs supports possible translation related functions. BC1 RNA evolved via a tRNA, most likely a $tRNA^{Ala}$ (Daniels and Deininger, 1985; Sakamoto and Okada, 1985; and work presented herein). BC200 RNA arose either via a monomeric Alu element (descendants of 7SL RNA (Ullu and Tschudi, 1984)) or directly via 7SL RNA (work presented herein). 7SL RNA is the structural component of the signal recognition particle (SRP) which is essential for protein translocation across the endoplasmic reticulum (Walter and Blobel, 1982). Both tRNAs and the SRP are associated with ribosomes during translation. Evidence that the binding sites for the 9 and 14 kilodalton protein subunits of the SRP are not only present in BC200 RNA but are also functional (J.-G. Cheng and J. Brosius and D. Zopf and P. Walter, UCSF, personal communication) and that the onset of BC1 RNA expression coincides with synaptogenesis further support this theory.

Interspersed repetitive DNA

A common feature of all eukaryotic genomes is the presence of repetitive DNA (Britten and Davidson, 1969, 1971). These repeated nucleotide sequences can be classified into two categories: tandemly repeated and interspersed repetitive DNA. Tandemly repeated sequences are generated by duplication at the DNA level. They encompass simple repeats of no greater than 6 base pairs in length, repeated greater than 10^6 times and can also include the genes of

multigene families (reviewed in Jelinek and Schmid, 1982). It is believed that the short sequence repeats are structural components of chromosomes, residing mostly at either the centromeres or telomeres (Jelinek and Schmid, 1982). The multigene families are unified by the cell's demand for large amounts of their gene products. This is exemplified by a number of different genes: tRNAs, 5S RNA and ribosomal RNA (reviewed in Long and Dawid, 1980), U1 and U2 snRNAs (Lund et al., 1984; Van Arsdell and Weiner, 1984) and the histones (reviewed in Kedes, 1979).

Interspersed repetitive DNA is classified into two categories based upon size: *l*ong (LINEs) and *s*hort (SINEs) *in*terspersed *e*lements (Singer, 1982). It is believed that the multiplication of both LINEs and SINEs is primarily driven by the amplification of only a small number of "master" or "source" genes (Deininger et al., 1992). The reverse transcription of RNA produced from these master genes, followed by the insertion of usually transcriptionally silent copies back into the genome via retroposition (Rogers, 1983; 1985; Van Arsdell, et al., 1981, Jagadeeswaran et al., 1981; Sharp, 1983) accounts for the amplification and dispersal of LINEs and SINEs (reviewed in Weiner et al., 1986). LINEs may even assist in the process of retroposition by encoding the necessary enzymes (Dombroski et al., 1991; Mathias et al., 1991). SINEs, which do not code for enzymes, require pre-existing cellular mechanisms for retroposition.

The prototype SINE is the human Alu sequence. The 500,000 Alu sequences dispersed thoughout the human genome comprise the most prodigious family of middle repetitive DNA, constituting approximately 5% of the genome (reviewed in Jelinek and Schmid, 1982). The earliest

Alu sequences arose from the 7SL RNA gene (Ullu and Weiner, 1984) while further amplification, as already stated, has been directed from a limited number of Alu "source" or "master" genes.

Primate Alu sequences are composed of two imperfectly duplicated repeat sequences separated by a variable length poly(A) stretch. Rodent Alu equivalents, the B1 family is monomeric and does not possess the primate duplicated structure. The primate Alu right monomer (Rm), which lacks 155 nucleotides of the 7SL RNA sequence, shows closest sequence homology to 7SL RNA. The left monomer (Lm), which is shorter by 31 base pairs, possesses an internal RNA polymerase III split promoter sequence which is capable of directing *in vitro* transcription. Models which attempt to explain the origin of the dimeric Alu sequences propose either the duplication of the 7SL derived Rm, with a subsequent deletion to produce the Lm, or the fusion of two monomers which arose independently from 7SL RNA genes (Jurka and Zuckerkandl, 1991; Quentin, 1992; reviewed in Weiner et al., 1986).

Rodents, unlike primates, have multiple SINE families. Only low levels of homology are shared between families (Rogers, 1985). The majority of these rodent SINE families arose from tRNA or tRNA genes and one of the most prodigious is the ID family of repeats. This element is found dispersed in the rat genome in 10^5 copies while the mouse, hamster and guinea pig genomes contain 10^4, 10^3 and 10^2 copies, respectively (Rogers, 1985; Deininger, 1989). Sequence comparison of these elements with various tRNAs (Daniels and Deininger, 1985; Sakamoto and Okada, 1985; and work presented herein) has revealed that the ID elements most likely arose via a $tRNA^{Ala}$.

Finding functions for interspersed repetitive DNA has proven elusive. Two extreme views have been formulated in this regard. One theory postulates that these sequences are involved in regulating gene expression (Britten and Davidson, 1969; Davidson and Britten, 1979). The other considers interspersed DNA "parasitic" and has characterized it as "selfish"(Dawkins, 1976; Orgel and Crick, 1980; Doolittle and Sapienza, 1980). As defined by Orgel and Crick (1980): "A piece of selfish DNA, in its purest form, has two distinct properties: (1) It arises when a DNA sequence spreads by forming additional copies of itself within the genome. (2) It makes no specific contribution to the phenotype."

Transcriptional regulation

Each cell is defined by the nature of its constituent molecules. The potentiality of the cell is determined by its genetic material while individuality is governed by the differential regulation of these genes. Transcriptional regulation is recognized to be one of the primary means of gene control for both eukaryotes and prokaryotes (Darnell, 1982). Understanding how this regulation is achieved is one of the main focuses of modern molecular and developmental biology. Three different RNA polymerases are utilized by eukaryotes to transcribe the approximately 10,000 different RNA species found per cell type. Each of these polymerases catalyzes the transcription of a different class of RNA.

RNA polymerase I

Polymerase I (pol I) synthesizes the 35-47S RNA precursor of the ribosomal 18S, 28S and 5.8S RNAs. Unique aspects of pol I include its nucleolar localization, production of only one transcript and that the promoters used in synthesizing this sole transcript are species-specific (reviewed in Sollner-Webb & Tower, 1986; Reeder, 1990; Sollner-Webb & Mougey, 1991). Necessary for the accurate and promoter-specific transcription are the species-specific promoter and at least two transcription factors, the upstream binding factor (UBF) and promoter selectivity factor (SL1) (Bell, et al., 1988; Learned et al., 1985, 1986). The highly conserved UBF binds to the promoter without need for additional factors (Bell at al, 1989; Pikaard et al., 1989) while SL1, the species-selectivity factor (Bell et al, 1989; Learned et al., 1985), forms a cooperative DNA binding complex with UBF, critical for transcriptional initiation (Bell et al., 1990). Regulation of rRNA transcription comes in response primarily to stimuli which alter rates of protein synthesis or cell growth (reviewed in Sollner-Webb & Tower, 1986). This may be effected via interactions directed towards the UBF-SL1 initiation complex (Comai et al, 1992).

RNA polymerase II

Polymerase II (pol II) transcribes all the genes coding for proteins and most of the genes that code for the snRNAs, which are involved in RNA processing. Pol II and a small group of "general" protein transcription factors are both necessary and sufficient for transcription initiation *in vitro* (reviewed in Kadonaga, 1990; Sawadogo and Sentenac,

1990; and Roeder, 1991). These general factors include TFIIA, TFIIB, TFIID (the only general factor with intrinsic site-specific DNA binding activity), TFIIE and TFIIF. However, recent studies indicate that the belief that all genes require these, or only these, same sets of general factors may be too simplified (Parvin et al, 1992). Regardless, the ordered interaction of these factors with pol II allows the formation of a functional preinitiation complex. The first step of this assembly, which promotes the sequential formation of the remainder of the complex and results in template commitment, is believed to be dependent upon the binding of TFIID, possibly aided by TFIIA, to the promoter via interaction with the TATA element (Sawadogo and Roeder, 1985; Van Dyke et al. 1988; Buratowski et al., 1989).

Genetic regulation in eukaryotes is primarily imposed at the level of transcription. This control, which may either stimulate or repress, is transmitted via the action of regulatory proteins that bind to specific sites in the promoters of pol II transcribed genes (reviewed in Maniatis *et al.*, 1987; Guarente, 1988; Mitchell and Tjian, 1989; Levine and Manley, 1989; Karin, 1990; Renkawitz, 1990). Numerous *cis*- and *trans*- acting factors have been identified (reviewed in Johnson and McKnight, 1989) and compiled (Locker & Buzard, 1990; Wingender, 1988; Jones *et al.*, 1988). The identification of these factors is allowing for the study of how transcription can be regulated to achieve differential gene expression in the context of developmental programs and intra- and extracellular cues. This may eventually lead to information on how these transcription factors are able to communicate with the transcriptional apparatus and are themselves regulated.

RNA polymerase III

RNA polymerase III is responsible for the synthesis of nuclear and cytoplasmic small RNAs (reviewed by Ciliberto *et al.*, 1983 and Geiduschek and Tocchini-Valentini, 1988). Examples include 5S RNA, tRNA, 7SL, 7SK, U6, certain virally encoded products such as the VAs I and II of adenovirus and EBERs I and II of the Epstein-Barr virus. Pol III is also responsible for the transcription of certain repetitive elements including the human and rodent Alu repetitive sequences and the rodent ID elements.

The promoters of pol III genes were the first eukaryotic regulatory regions to be examined. Surprisingly, unlike the well studied prokaryotic examples, the first eukaryotic promoter, that of the 5S ribosomal RNA gene of *Xenopus borealis*, was found within the coding region of the gene (Sakonju et al., 1980; Bogenhagen et al., 1980). This result was soon corroborated by the isolation of the first eukaryotic transcription factor. It bound to the internal region of cloned *Xenopus* 5S genes and was necessary for their accurate transcription (Engelke et al., 1980). The sequence elements necessary and sufficient for transcription of tRNA genes (Kressman et al., 1979; De Franco et al., 1980; Galli et al., 1981; Hofstetter et al., 1981) and the adenovirus VAI gene (Fowlkes & Shenk, 1980) were also found to be intragenic.

These intragenic promoters were further delineated. Two domains, boxes A and B, which are highly conserved amongst the tRNA-like genes and possess several invariant nucleotides (Sharp et al., 1984) were defined by deletion and substitution experiments (Hofstetter *et al.*, 1981; Galli *et al.*, 1981; and Sharp *et al.*, 1981). Site directed mutagenesis

of the 5S RNA gene's intragenic promoter revealed a similar discontinuous promoter structure (Geiduschek and Tocchini-Valentini, 1988). The 5' region of the 5S RNA gene internal control region is both structurally and functionally equivalent to the tRNA box A (Ciliberto et al., 1983) while the 3' region is not related to box B, but unique to 5S genes, and is named box C.

The chromatographic fractionation of cellular extracts has revealed that in addition to pol III at least two common protein factors are required for the transcription of all class III genes, TFIIIB and TFIIIC. An additional factor, TFIIIA, is required for the transcription of 5S RNA (Segall et al., 1980; Shastry et al., 1982). TFIIIA's binding is strongly influenced by box C (Sakonju and Brown, 1982); interestingly, TFIIIA can also bind 5S RNA (Pelham and Brown, 1980; Honda and Roeder, 1980). TFIIIC's binding is strongly influenced by box B and once bound, TFIIIC cooperatively interacts with TFIIIB, which is itself unable to bind to DNA directly, to correctly position it upstream of the transcription start site (Lassar et al., 1983; Bieker et al., 1985). In yeast, once TFIIIB has formed a stable complex it is resistant to removal by high salt concentrations and heparin, conditions which dissociate TFIIIA and TFIIIC, and is able to direct multiple rounds of transcription (Kassavetis et al., 1989 and 1990).

Following the *in vitro* characterization of the intragenic pol III promoters, it became apparent that important regulatory elements can also exist in the 5' flanking regions of pol III genes. These essential 5' promoters have been demonstrated for a number of genes (Sprague et al., 1980; Dingermann et al., 1982; Hipskind and Clarkson, 1983; Johnson and Raymond, 1984; Ullu and Wiener, 1985; Selker et al., 1986; Morry and

Harding, 1986; Garcia et al.,1987; Howe and Shu, 1989). The most extreme examples of upstream pol III promoters are found in the 7SK and U6 genes. Neither gene possesses a consensus box B or C element. While a consensus box A sequence is present, it is not required for transcriptional activity. Instead, upstream promoter elements resembling those for pol II transcribed genes, in both sequence and location, are necessary and sufficient for transcription (Murphy et al., 1987; Carbon et al., 1987; Das et al., 1988; Kleinert and Benecke, 1988; Kunkel and Pederson, 1988). The functional sequences shared by these two pol III transcribed genes are also found with pol II transcribed genes and include a TATA sequence, a proximal sequence element (found in the snRNA pol II genes) and distal octamer motifs (reviewed in Murphy *et al.*, 1989).

Not only do the genes transcribed by pol III use many of the same strategies for transcription initiation complex formation as do the genes transcribed by pol II, but pol III transcribed genes may also use the same pol II components (reviewed in Murphy et al., 1989 and Gabrielsen and Sentenac, 1991). Several groups have now shown that the TATA-binding protein is involved in the *in vitro* transcription of class III genes (Lobo et al., 1991; Margottin et al., 1991; and Simmen et al., 1991). Another example of the interconnectedness of pol II and III transcription has been described wherein the polymerase specificity of several U RNA genes has been switched (Mattaj *et al.*, 1988; Lobo and Hernandez, 1989; Waibel and Filipowicz, 1990).

Transgenic mice

One of the most powerful methodologies currently available for studying gene regulation *in vivo* is the use of transgenic mice (reviewed in Palmiter and Brinster, 1986; Hanahan, 1988). Utilizing these animals, one can produce families of mice carrying exogenously added new or altered genes stably integrated within their genomes. Study of the control and tissue specificity of gene regulation were the first uses of this technique (Gordon et al., 1980; Wagner et al., 1981; Harbers et al., 1981; Costantini and Lacy, 1981; Brinster et al., 1983), and it still remains one of the foremost techniques for studying tissue specific transcriptional regulation *in vivo*.

The procedure for generating transgenic mice involves microinjection of a DNA solution containing the gene of interest into the pronuclei of a fertilized egg (one cell stage), then re-implanting this egg, which now carries the "transgene", into the oviduct of a pseudopregnant female mouse. Approximately 10 - 25% of the mice born from the injected embryos will carry one or multiple copies of the transgene. Each "transgenic" mouse is unique with respect to chromosomal site of integration and transgene copy number. Mating of these transgenics, heterozygotic with respect to the transgene, with a non-transgenic mouse, results in the progeny receiving the transgene with the expected Mendelian inheritance pattern. Within each lineage, the mice tend to be genetically similar. Breeding within lineages is performed to achieve transgene homozygotes.

That correct tissue-specific expression patterns could be duplicated in transgenic mice was originally demonstrated with the immunoglobulin κ gene (Brinster et al., 1983) and for numerous genes

thereafter (Palmiter and Brinster, 1986). However, transgenic mice within the same lineage tend to be quite similar with respect to transgene expression while lineages with different insertions of the same transgene can display marked variation in expression. This effect has been attributed to integration of the transgene at different chromosomal locations (Adams et al., 1987). By establishing the correct "chromosomal milieu", positional effects can be overcome. An example is provided by the human β globin gene (Grosveld et al., 1987). By providing 50 kb of 5' and 20 kb of 3' flanking sequence possessing DNAse hypersensitive sites high levels of tissue specific expression, regardless of the integration site, were achieved.

DNA Sequencing

In 1977 two new methods for the sequencing of DNA were reported by Maxam and Gilbert (1977) and Sanger et al., (1977). The method of Maxam and Gilbert involves chemical modifications while that of Sanger relies on enzymatic means. However different, the basic principles are the same:

1. fragments of DNA are isolated;
2. a set of standard reactions are applied to generate four sets of labelled molecules, each with base specific ends;
3. these four mixtures are separated on a mass basis (proportional to fragment length) by polyacrylamide gel electrophoresis; and,

4. the resulting patterns are interpreted to reveal the original sequence.

The key objective of both methodologies is to generate a population of oligonucleotides which keep one end fixed and the other end terminating at one of the four bases. It is at this step where the two techniques differ. Although both are reliable the Sanger method is faster and uses fewer hazardous materials making it ideal for large-scale tasks.

The Sanger technique enzymatically synthesizes a mixture of DNA fragments complementary to the original single-stranded DNA. The mixture of fragments all share the same 5' sequence but differ in their point of termination. Synthesis is achieved by the use of DNA polymerase I. The template directed enzyme, DNA polymerase I, has three important features for sequencing purposes. First, it can only begin synthesis from a double-stranded primer. Second, it adds deoxynucleotides to the existing 3' hydroxyl terminus of that double-stranded DNA primer. Third, DNA polymerase can incorporate base analogs. When a 2'3'-dideoxy (dd) analog of one of the four deoxynucleotides is incorporated into a nascent strand, further elongation of that chain is blocked: chain-terminated fragments are created. By specifically using only one of each dideoxyanalog (ddTTP, ddCTP, ddGTP, or ddATP) in each reaction mixture, one is selecting for termination at that particular base.

For large scale sequencing projects three strategies have been adopted. The first of these is termed "shotgun DNA sequencing". It refers to the spanning of the region of interest with random DNA subfragments and using these to rapidly accumulate nucleotide

sequence data. These randomized fragments are then ordered and assembled to generate the full-length sequence. Another approach was originally described by Henikoff (1984) and involves generating unidirectional deletion mutations in the subcloned DNA fragment. Deletion mutants are prepared by cleaving the sequence at a known point in the vector's polylinker region using two restriction enzymes, one of which leaves a 3'-overhanging end and the other either a 5'-overhanging end or a blunt end. The 5'-overhanging end is selected to be contiguous to the DNA insert to be sequenced. Exonuclease III is able to catalyze the stepwise removal of nucleotides in a 3' to 5' direction from 5' overhanging or blunt ends. By aliquoting the linearized DNA sample and allowing the nuclease defined amounts of time to digest the DNA, deletions of varying length are generated. These nested deletions are then treated with S1 nuclease to remove the single-stranded DNA and the molecule self-ligated. In this manner, the primer annealing point remains constant while shortened versions of insert DNA have been enzymatically divided into smaller, overlapping subclones.

The final method uses the sequence generated from each sequencing run to define a new sequencing primer. This primer site is located at increasingly further points from the original start site. In this manner one is able to "walk" across the DNA insert. This method with its continual need for new oligonucleotide primers is limited only by the expense of generating a continuous stream of synthetic oligonucleotides and the time required to synthesize and process them.

A new concept developed by Smith et al. (1986) replaces the need for radioactive labels by the use of four different fluorophores covalently attached to the oligonucleotide primer. This allows for the continuous

electrophoresis of the newly synthesized, fluorophore specific, chain-terminated strands. As they migrate past the bottom of a polyacrylamide gel tube, the fluorescent fragments are detected by a combination of lasers with filters for detection at specific wavelengths. Sequence information can be directly conveyed to a computer.

CHAPTER 2

MATERIALS AND METHODS

Materials

The various λ phage vectors containing rat, mouse, Syrian and Chinese hamster and guinea pig BC1 RNA gene isolates were provided by K Eisinger, S. Klco, S. Friedman and S. Kaplan. E. coli strain TG1 was a kind gift of T. Gibson (MRC, Cambridge). Bacterial media 2xTY contained 10 g/l bacto-tryptone, 10 g/l yeast extract and 5 g/l of NaCl. Restriction enzymes, exonuclease III, T4 DNA ligase, and Nuclease Bal 31 were products of New England Biolabs and enzyme buffers were prepared as directed by the supplier. Klenow DNA polymerase was purchased from Boehringer Mannheim. Oligonucleotides were purchased from the Mt. Sinai DNA Core Facility, dideoxynucleotides from Pharmacia and all radioactive nucleotides from DuPont-New England Nuclear.

The composition of other solutions used were as follows:

10x T.B.E. (/L) :	108 g Tris base
	55 g boric acid
	9.3 g EDTA(Na_2)

T.B.E dye mix (/100ml) :	10 ml 10x T.B.E.
	0.1 g bromophenol blue
	20 g sucrose
T.E. :	10 mM Tris (pH 8.0-8.5)
	.1 mM EDTA(Na_2)
10x ligase buffer :	500 mM Tris (pH 7.5)
	100 mM $MgCl_2$
	100 mM DTT
20% PEG (/100ml) :	20 g polyethylene glycol (M.W. 8000)
	14.6 g NaCl
TFB :	10 mM MES
	100 mM $RbCl_2$
	45 mM $MnCl_2 \cdot 4H_2O$
	10 mM $CaCl_2 \cdot 2H_2O$
	3 mM Hexaminecobaltic chloride
DTT/KAc :	2.25 M DTT
	40 mM Potassium acetate (pH 6.0)
T.M. :	100 mM Tris (pH 8.5)
	50 mM $MgCl_2$

NTP mixes :

nucleotide mix:	T	C	G	A
0.5 mM dTTP	25	500	500	500 µl
0.5 mM dCTP	500	25	500	500
0.5 mM dz^{7}GTP	500	500	25	500
10 mM ddTTP	50	0	0	0
10 mM ddCTP	0	8	0	0
10 mM ddGTP	0	0	16	0
10 mM ddATP	0	0	0	1
T. E.	1000	1000	1000	1000

Formamide dye mix : 100 ml Formamide (deionised with mixed bed resin)
0.1 g xylene cyanol
0.1 g bromophenol blue
2 ml 0.5M EDTA (ph 8.0)

40% Acrylamide (/L) : 380 g acrylamide
20 g NN-methylenebisacrylamide
Bring volume to one liter with deionised water and when dissolved stir with 20 g Amberlite MB 1 then filter through sintered glass.

0.5x TBE 6% gel mix : 75ml 40% acrylamide
25ml 10x TBE
230 g urea (BRL ultrapure or USB ultrapure)
Bring volume to 500 ml with deionised water and filter through sintered glass.

5.0x TBE 6% gel mix 30 ml 40% acrylamide
100 ml 10x TBE
92 g urea
10 mg bromophenol blue
Bring volue to 200 ml with deionised water and filter through sintered glass.

Methods

DNA sequencing and analysis.

Template preparation

From a streaked plate of TG1 cells grown on minimal media a single colony was picked and 10 ml of 2xYT was inoculated and grown with shaking at 37° C overnight. Three hundred µl of this overnight culture was added to 30 ml of un-inoculated 2xYT and grown with shaking at 37° C for two hours. After centrifugation at 2000 RPM for 10 minutes at 4° C the 2xYT was poured off and the pellet drained. For

transformation, 3 mls of cold TFB was added to the cells, resuspended gently and left on ice for 10 minutes. Then 100 µl of DMSO was added and the cells left on ice for 5 minutes - then 100 µl of DTT/KAc and the cells were left on ice for 10 minutes - then 100 µl of DMSO and the cells were left on ice for 5 minutes. Each time a reagent was added, the cells were resuspended gently. After the last addition of DMSO and five minutes had elapsed, 200 µl of the competent cells was added to the a determined amount of DNA. This was incubated on ice for 45 minutes. At the end of this time, the cells were heat shocked at 42° C for 1 minute, placed back on ice for 2 minutes and then 200 µl of 2xYT is added to the tube. This is then grown with shaking for 45 minutes at 37° C. Cells are then plated onto agar plates containing the required antibiotics.

Single-stranded templates were prepared by a modification of pre-existing methods. Namely, isolated colonies from O/N plates were innoculated into 1.5 µls of 2 x YT media to which ampicillin (100 µg/µl) and 5 µl of M13KO7 (10^9 pfu/µl) had been added. These were then grown with shaking (300rpm) in a 37° C air incubator for 7-8 hours. Isolation of single stranded template was then as per Bankier *et. al.*, (1988). After growth, the culture was transfered to a 1.5 µl microfuge tube and centrifuged for 5 minutes. The supernatant was carefully transfered to another microfuge tube and 200 µl of 20% PEG was added and then the tube vortexed and incubated at room temperature for fifteen minutes. The tube was then centrifuged for five minutes, the PEG removed by aspiration and the tube respun to remove any residual PEG. After resuspending in 100 µl of TE, 50 µl of buffer saturated phenol was added and the mixture vortexed. After 5 minutes the mixture was vortexed again and spun for 3 minutes in a microfuge. To the upper aqueous

layer 1/10 volume of 3M sodium acetate and 2.5 volumes of 95% ethanol was added and precipitated overnight at -20 degrees C.

The ethanol was poured off following 10 minutes centrifugation in the microfuge. One ml of 95% ethanol was added, centrifuged for 5 minutes and then poured off. The remaining pellet was vacuum dried for 5 minutes and then redissolved in 80 µl TE and stored at -20 degrees C.

Sequencing reactions

The reactions described were performed in Falcon 3911 96 well U bottom microtitre plates (Becton Dickinson Co.) and the 2 µl additions were performed using a Hamilton PB600 repetitive dispenser fitted with a 1710 LT syringe (Hamilton). To each of four wells marked "T", "C", "G", "A", 2 µl of primer/TM mix was added and against the opposite well wall 2 µl of template DNA was added. The tray was covered with Saran wrap and centrifuged briefly (maximum speed of 2000 RPM) and oven incubated at 55° C for 45 minutes. The tray was re-centrifuged to concentrate as much of the condensation as possible and then 2 µl of the appropriate dideoxynucleotide mix was added into each well. Klenow fragment DNA polymerase I was added to the diluted [α-35S]dATP mix and immediately 2 µl of dNTP chase solution was added and again briefly centrifuged and incubated at room temperature for 15 minutes. At this point the mixes were frozen at -20 ° C or if used immediately, 2 µl of formamide dye mix was added to each well and centrifuged briefly as above. After heat denaturing the reactions in an oven at 80 ° C for 15 minutes to reactions were loaded onto gels and electrophoresed.

Data analysis

The data generated were manually entered from autoradiographs using the Staden sequence analysis programs on a VAX 6310 system, running the VMS 5.4 operating system, available through the Columbia University Comprehensive Cancer Center Computing Facility. Analysis was performed using the GCG suite of programs (version 7.0) [J. Devereux et al., 1984) made available by the Department of Biomathematics, Mount Sinai School of Medicine.

Guinea pig Southern and Northern blot analyses

Guinea pig DNA (10 μg) was digested with a series of different restriction enzymes as per the manufacturer's recommendations. The resultant digests were phenol/chlorofom extracted and ethanol precipitated and then redissolved in TE (10 mM:1 mM). The samples were then electrophoresed on a 1% agarose gel and then transferred overnight onto a nylon membrane (GeneScreen, DuPont-NEN) membrane and immobilized by UV crosslinking (Stratalinker, Stratagene). The filter was prehybridized for 3 hours at 50° C in a solution containing 5x Denhardt's solution (1x= 0.02% Ficoll 400/ 0.02% polyvinvlypyrrolidone/ 0.2% bovine serum albumin), 5x SSC, 0.05% SDS, 50% deionized formamide, and 100 μg of denatured salmon sperm DNA per ml. A PCR generated probe containing (^{32}P) dCTP, directed between upstream primers positioned from the end of the ID sequence to the unique 3' end. After the addition of 1 x 10^7 cpm of probe and hybridization for 12 hours, the filter was washed 3 times with 1 x SSC,

0.1% SDS at 55° C for 30 minutes each. A fourth wash with 0.5 x SSC, 0.1% SDS at 55° C for 30 minutes was then performed. The filter was then exposed, with intensifier screen, overnight at -80° C, onto Fuji X-ray film.

For the guinea pig Northern, RNA samples from three guinea pig tissues were electophoresed on a denaturing 6% polyacrylamide 7M urea gel using 30 µg of brain, liver and kidney RNA. The samples were then transferred and fixed onto a nylon filter. A 33-mer (5'AAAGGTTGTTTGTGTGCGCAGTTACCTTGTTTG-3') complementary to the guinea pig's unique 3' sequence was 5' end-labeled with (^{32}P)ATP (Maniatis et al., 1982). Prehybridzation, hybridization and washing conditions were as noted above except that the deionized formamide concentration was decreased to 20%, the temperature used was 50° C, instead of 55° C , and the final wash used 1 x SSC, 0.1 SDS %. The filter was exposed to film at -80° C for 24 hours.

PCR labeling of oligonucleotides

Probes used for Southern analyses of transgenic samples were labeled to high specific activity using the following protocol kindly provided by K. "papa" Kelley :

1. The labeling reaction is prepared in a 400µl Eppendorf tube on ice as follows:

5 µl 10x buffer

5 µl dNTP mix (- dCTP)

5 µl oligonucleotide mix

5 µl 100 mM β-mercaptoethanol

1 µl DNA template

27 µl 32^P dCTP (3000 Ci/mmol; 10µCi/µl)

2 µl diluted Taq polymerase (Cetus)

2. The PCR reaction is overlayed with 30 µl of light mineral oil, and amplified in a thermal cycler:

1. Ramp to 94° C @ 2 sec./° C
2. 94° C x 4 min.
3. 94° C x 30 sec.
4. Ramp to 55° C @ 2 sec/ ° C
5. 55° C x 1 min.
6. Ramp to 72° C @ 2 sec/° C
7. 72° C x 4 min.
8. Repeat lines 3-7 for 30 cycles
9. 72° C x 10 min.
10. Ramp to 15° C @ 25 sec/° C
11. Cool to 15° C

3. Free nucleotides are removed and the probe is used immediately.

Notes on buffers: 10x buffer : 166 mM NH_4SO_4, 67 mM Tris (pH 8.8), 2.0 mM MgCl, 0.2 mg/ml BSA. Oligonculeotide mix: contains 0.1 ug/µl of each nucleotide. dNTP mix: 2.5 mM of dATP, dGTP and dTTP.

CHAPTER 3

RESULTS

Organization, structure and sequence analysis of the BC1 RNA gene of rodents

BC1 RNA was originally detected in rat brains with a probe containing the ID element (Sutcliffe et al., 1982). As previously described (DeChiara and Brosius, 1987), BC1 RNA possesses a tripartite structure: a 5' ID containing portion, a central A-rich region and a terminal 25 nucleotide (nt) region, which is unrelated to the ID region, termed BC1 unique (Figure 1). This unique sequence, when used at high stringency as a probe for Southern blots of rat genomic DNA reveals the presence of a single hybridizing band (not shown), suggesting that BC1 RNA is the product of a single gene.

To isolate the BC1 RNA gene from rodents, the unique 3' regions from a number of different rodent species were determined by isolating and sequencing the BC1 cDNA sequences from these rodent species (K. Eisinger, S. Klco, S. Friedman, and S. Kaplan). Based on the sequence of the cDNAs isolated, oligonucleotide probes were designed to the specific BC1 unique regions of each rodent sequence. These were then used to screen λ genomic libraries of the respective species. The genes isolated and sequenced are shown schematically in figure 2 and the sequences are presented in the Appendix.

Figure 2. Diagram of the rodent BC1 RNA genes and a BC1 RNA pseudogene. The structures represent the sequences presented in the Appendix and are drawn to scale. The figures are aligned to reveal homologous sequences and are centered about the BC1 RNA coding region, represented by an oval. The presence of B1 repetitive elements is highlighted by overlapping circles and their orientation, with respect to their internal promoter sequences, is defined by the direction of the arrows. Positions of the transcribed regions, B1 elements, and non-homologous regions (filled-in boxes) are indicated. The Syrian hamster sequence, a BC1 RNA pseudogene, does not contain non-B1 containing flanking sequences homologous to any of the other sequences.

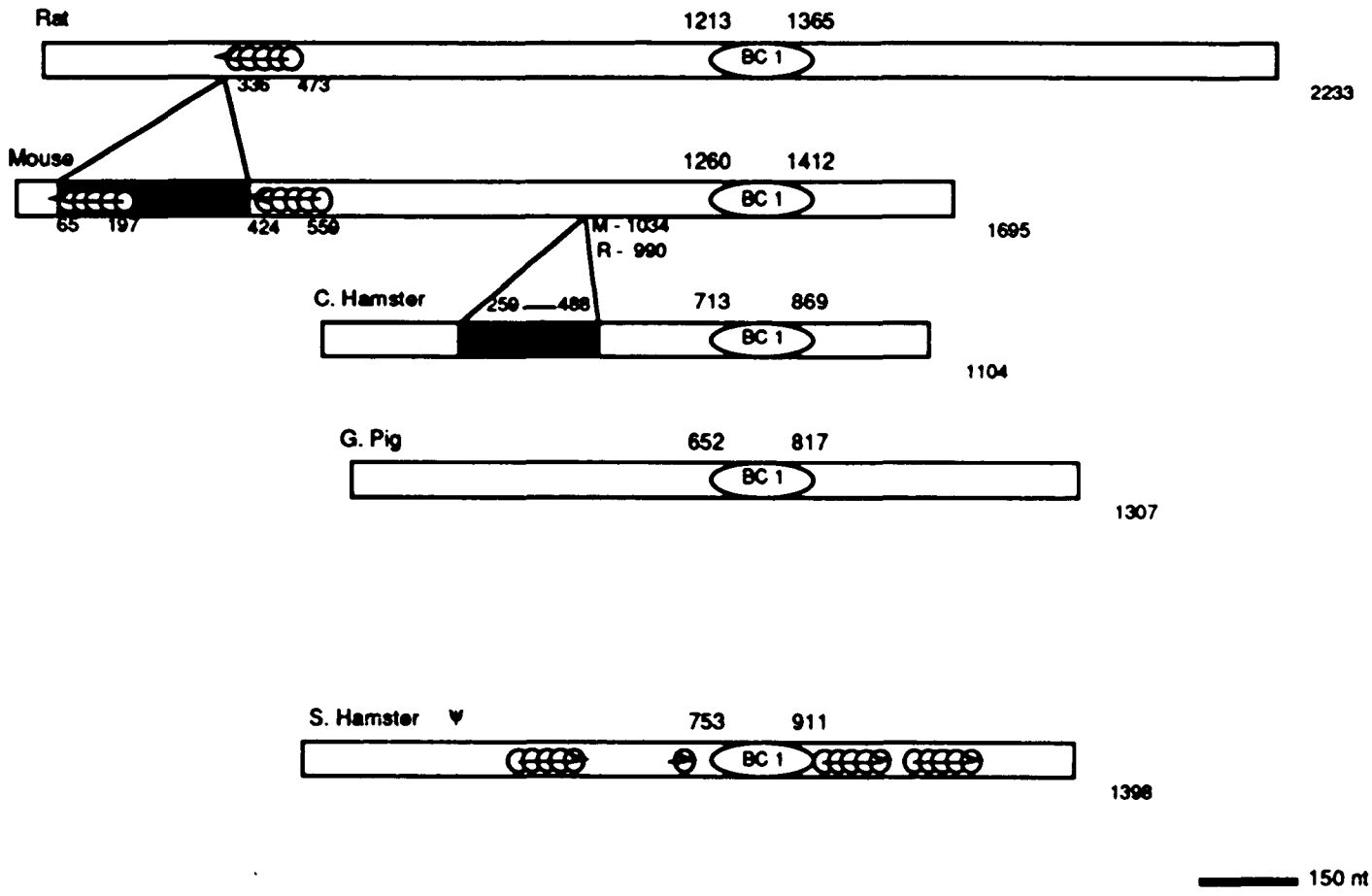

Rat
1213
1365
BC 1
336
473
2233
Mouse
1260
1412
BC 1
65
197
424
559
M - 1034
R - 990
1695
C. Hamster
259
488
713
869
BC 1
1104
G. Pig
652
817
BC 1
1307
S. Hamster
753
911
BC 1
1398
150 nt

Figure 3 summarizes the pairwise comparisons of the rodent BC1 RNA gene and flanking sequences from the rodent sequences analyzed. As shown in figure 4, the BC1 RNA coding region sequences between rat, mouse and chinese hamster, all Old World rodents, share greater than 98% sequence homology. The guinea pig, a New World rodent, sequence shows 87% homology with Chinese hamster and 89% homology with the rat and mouse sequences. What may be of particular relevance is that for all species the regions resembling the intragenic promoter sequences of tRNAs are conserved (underlined in the rat sequence).

The 5' and 3' flanking sequences among rat, mouse and hamster genes are also highly conserved (Figures 5 and 6). In more distant areas conservation of sequence drops to levels consistent with neutral drift relative to the evolutionary distances of the species involved. The relatively high sequence homology seen in these flanking areas may be explained by evolutionary pressure to maintain transcriptional control elements, which have been shown to be necessary in the *in vitro* transcription of the rat BC1 RNA gene (Chapter 5). These elements resemble RNA polymerase II promoters. They include a TATA box, a PSE element and octamer motifs.

The guinea pig flanking regions are not as well conserved. However, the 5' flanking region of the guinea pig gene does maintain the TATA box and elements of the PSE sequence which have been shown to be involved in regulating BC1 RNA transcription *in vitro*. The high degree of divergence beyond these sequences may be attributable to a fast evolutionary clock (with its increased rate of nucleotide substitutions) or to a greater than acknowledged separation between the origin of these

Figure 3. Summary of the pairwise comparisons of the rodent BC1 RNA gene and flanking sequences. The sequence from the four rodent species was divided into the BC1 coding region and into two blocks of up- and downstream flanking sequences, each consisting of 100 base pairs.

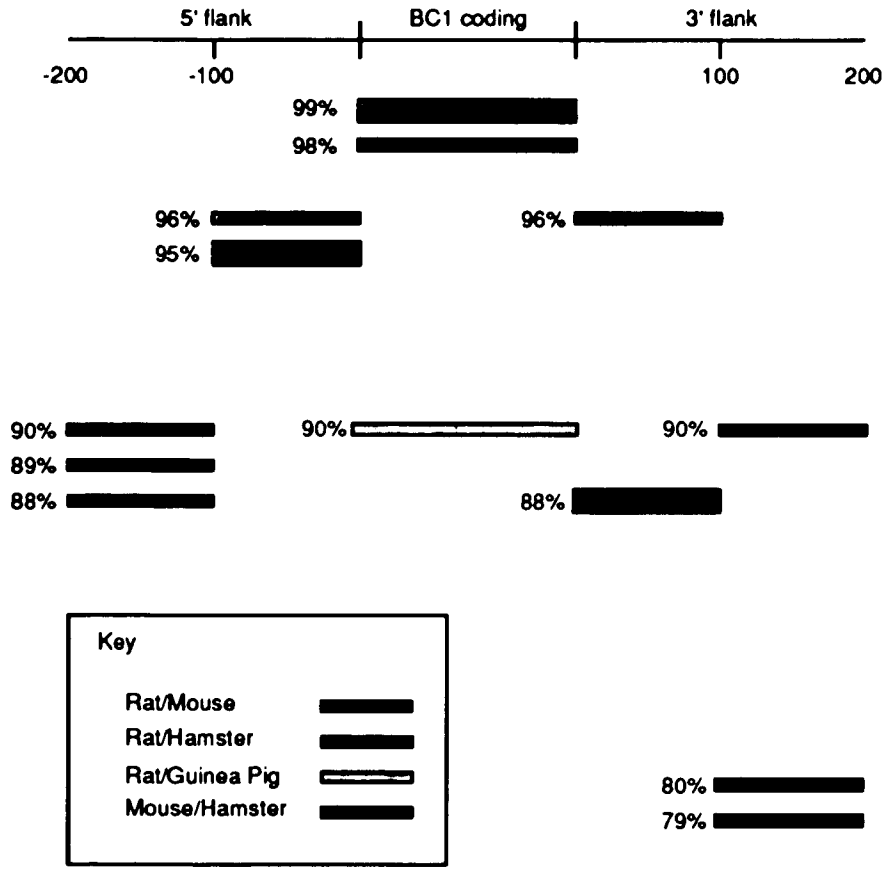

5' flank
BC1 coding
3' flank
-200
-100
100
200
99%
98%
96%
95%
96%
90%
89%
88%
90%
90%
88%
Key
Rat/Mouse
Rat/Hamster
Rat/Guinea Pig
Mouse/Hamster
80%
79%

Figure 4. Alignment of the BC1 RNA coding region of several rodent species. The length of the pairwise alignment and the percent homology is indicated between the compared sequences. As with all the alignments shown, vertical lines, "|", represent identical nucleotides and dots ".", represent spaces introduced to maximize the alignment. Analysis was initially performed using the program BESTFIT contained within the GCG suite of programs (Devereux et al., 1984), made available by the Department of Biomathematics, Mount Sinai School of Medicine, and then further refined manually.

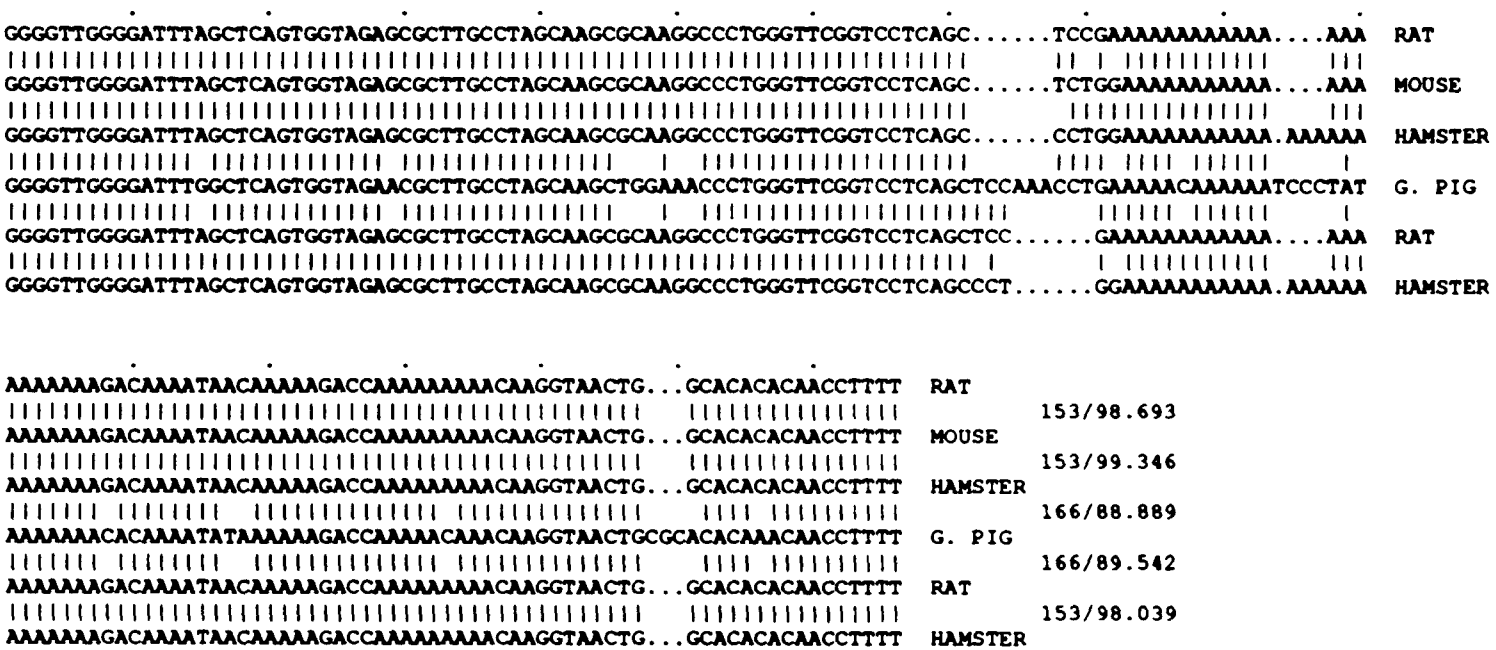

GGGGTTGGGGATTTAGCTCAGTGGTAGAGCGCTTGCCTAGCAAGCGCAAGGCCCTGGGTTCGGTCCTCAGC......TCCGAAAAAAAAAAAA....AAA RAT
GGGGTTGGGGATTTAGCTCAGTGGTAGAGCGCTTGCCTAGCAAGCGCAAGGCCCTGGGTTCGGTCCTCAGC......TCTGGAAAAAAAAAAA....AAA MOUSE
GGGGTTGGGGATTTAGCTCAGTGGTAGAGCGCTTGCCTAGCAAGCGCAAGGCCCTGGGTTCGGTCCTCAGC......CCTGGAAAAAAAAAAA.AAAAAA HAMSTER
GGGGTTGGGGATTTGGCTCAGTGGTAGAACGCTTGCCTAGCAAGCTGGAAACCCTGGGTTCGGTCCTCAGCTCCAAACCTGAAAAACAAAAAATCCCTAT G. PIG
GGGGTTGGGGATTTAGCTCAGTGGTAGAGCGCTTGCCTAGCAAGCGCAAGGCCCTGGGTTCGGTCCTCAGCTCC......GAAAAAAAAAAA....AAA RAT
GGGGTTGGGGATTTAGCTCAGTGGTAGAGCGCTTGCCTAGCAAGCGCAAGGCCCTGGGTTCGGTCCTCAGCCCT......GGAAAAAAAAAA.AAAAAA HAMSTER
AAAAAAAGACAAAATAACAAAAAGACCAAAAAAAAACAAGGTAACTG...GCACACACAACCTTTT RAT
153/98.693
AAAAAAAGACAAAATAACAAAAAGACCAAAAAAAAACAAGGTAACTG...GCACACACAACCTTTT MOUSE
153/99.346
AAAAAAAGACAAAATAACAAAAAGACCAAAAAAAAACAAGGTAACTG...GCACACACAACCTTTT HAMSTER
166/88.889
AAAAAAACACAAAATATAAAAAAGACCAAAAACAAACAAGGTAACTGCGCACACAAACAACCTTTT G. PIG
166/89.542
AAAAAAAGACAAAATAACAAAAAGACCAAAAAAAAACAAGGTAACTG...GCACACACAACCTTTT RAT
153/98.039
AAAAAAAGACAAAATAACAAAAAGACCAAAAAAAAACAAGGTAACTG...GCACACACAACCTTTT HAMSTER

Figure 5. The 5' flanking regions of the BC1 RNA gene of three rodent species. The 200 nucleotides directly upstream of the start site are shown.

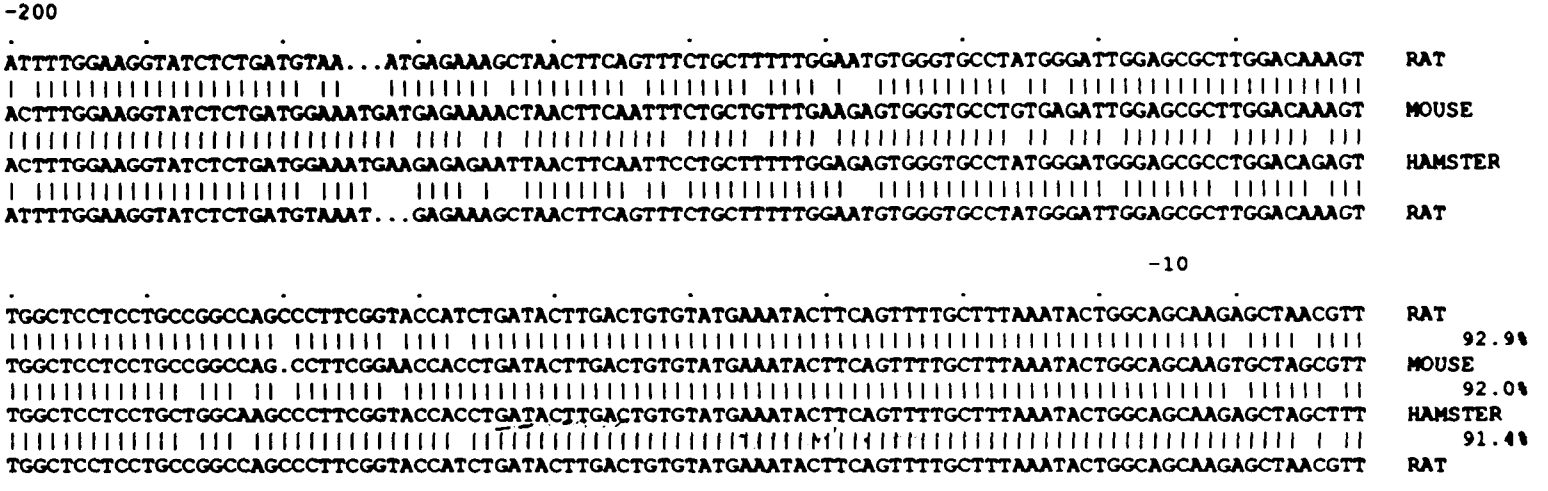
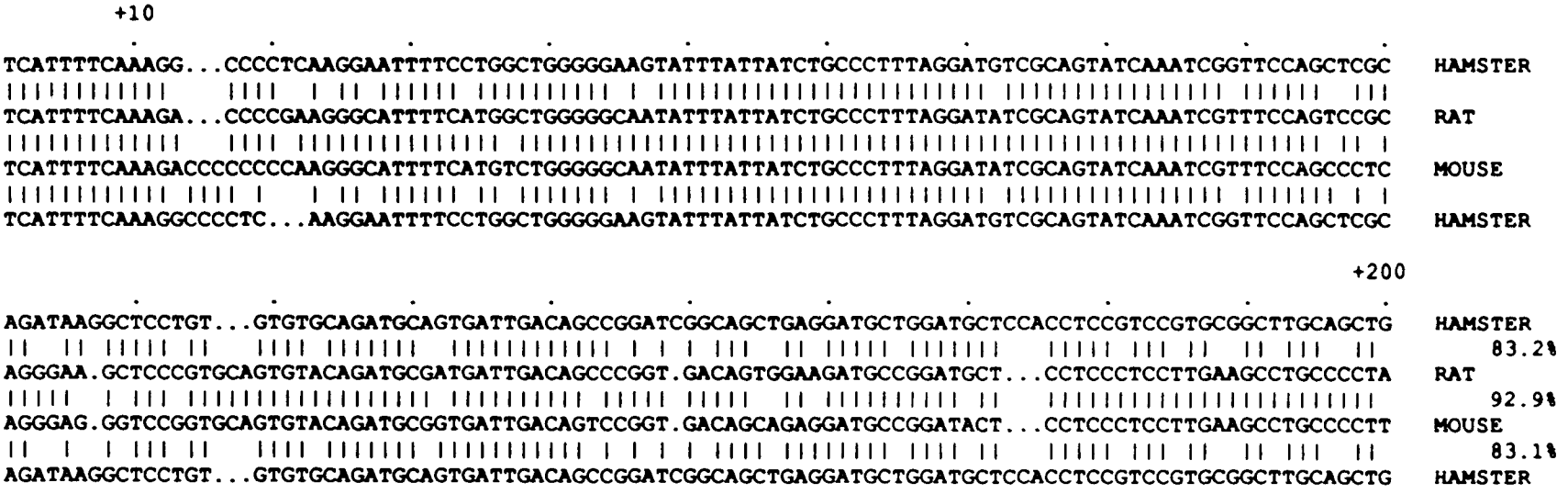

-200
ATTTTGGAAGGTATCTCTGATGTAA...ATGAGAAAGCTAACTTCAGTTTCTGCTTTTTGGAATGTGGGTGCCTATGGGATTGGAGCGCTTGGACAAAGT
RAT
ACTTTGGAAGGTATCTCTGATGGAAATGATGAGAAAACTAACTTCAATTTCTGCTGTTTGAAGAGTGGGTGCCTGTGAGATTGGAGCGCTTGGACAAAGT
MOUSE
ACTTTGGAAGGTATCTCTGATGGAAATGAAGAGAGAATTAACTTCAATTCCTGCTTTTTGGAGAGTGGGTGCCTATGGGATGGGAGCGCCTGGACAGAGT
HAMSTER
ATTTTGGAAGGTATCTCTGATGTAAAT...GAGAAAGCTAACTTCAGTTTCTGCTTTTTGGAATGTGGGTGCCTATGGGATTGGAGCGCTTGGACAAAGT
RAT
-10
TGGCTCCTCCTGCCGGCCAGCCCTTCGGTACCATCTGATACTTGACTGTGTATGAAATACTTCAGTTTTGCTTTAAATACTGGCAGCAAGAGCTAACGTT
RAT
92.9%
TGGCTCCTCCTGCCGGCCAG.CCTTCGGAACCACCTGATACTTGACTGTGTATGAAATACTTCAGTTTTGCTTTAAATACTGGCAGCAAGTGCTAGCGTT
MOUSE
92.0%
TGGCTCCTCCTGCTGGCAAGCCCTTCGGTACCACCTGATACTTGACTGTGTATGAAATACTTCAGTTTTGCTTTAAATACTGGCAGCAAGAGCTAGCTTT
HAMSTER
91.4%
TGGCTCCTCCTGCCGGCCAGCCCTTCGGTACCATCTGATACTTGACTGTGTATGAAATACTTCAGTTTTGCTTTAAATACTGGCAGCAAGAGCTAACGTT
RAT

Figure 6. The 3' flanking regions of the BC1 RNA gene of three rodent species. The 200 nucleotides directly flanking the termination site are shown starting at position +1.


+10
TCATTTTCAAAGG...CCCCTCAAGGAATTTTCCTGGCTGGGGGAAGTATTTATTATCTGCCCTTTAGGATGTCGCAGTATCAAATCGGTTCCAGCTCGC HAMSTER
TCATTTTCAAAGA...CCCCGAAGGGCATTTTCATGGCTGGGGGCAATATTTATTATCTGCCCTTTAGGATATCGCAGTATCAAATCGTTTCCAGTCCGC RAT
TCATTTTCAAAGACCCCCCCCCAAGGGCATTTTCATGTCTGGGGGCAATATTTATTATCTGCCCTTTAGGATATCGCAGTATCAAATCGTTTCCAGCCCTC MOUSE
TCATTTTCAAAGGCCCCTC...AAGGAATTTTCCTGGCTGGGGGAAGTATTTATTATCTGCCCTTTAGGATGTCGCAGTATCAAATCGGTTCCAGCTCGC HAMSTER
+200
AGATAAGGCTCCTGT...GTGTGCAGATGCAGTGATTGACAGCCGGATCGGCAGCTGAGGATGCTGGATGCTCCACCTCCGTCCGTGCGGCTTGCAGCTG HAMSTER
83.2%
AGGGAA.GCTCCCGTGCAGTGTACAGATGCGATGATTGACAGCCCGGT.GACAGTGGAAGATGCCGGATGCT...CCTCCCTCCTTGAAGCCTGCCCCTA RAT
92.9%
AGGGAG.GGTCCGGTGCAGTGTACAGATGCGGTGATTGACAGTCCGGT.GACAGCAGAGGATGCCGGATACT...CCTCCCTCCTTGAAGCCTGCCCCTT MOUSE
83.1%
AGATAAGGCTCCTGT...GTGTGCAGATGCAGTGATTGACAGCCGGATCGGCAGCTGAGGATGCTGGATGCTCCACCTCCGTCCGTGCGGCTTGCAGCTG HAMSTER

rodent suborders.

Neural specific expression of BC1 RNA has been conserved for 55 million years

Of the rodent sequences examined, the guinea pig BC1 RNA gene shows the greatest degree of sequence divergence. Graur et al., (1991) have recently suggested that guinea pigs may not be rodents: they may define a separate order of mammals. This suggestion was made to attempt to explain the "molecular paradox" that most available protein and nucleic acid sequences from guinea pig exhibit an unusual level of dissimilarity when compared to their myomorph homologues, e.g. rats, mice and hamsters. To assess whether the BC1 RNA gene is also present as a single copy gene in guinea pig, as in the myomorph rodents, a Southern analyses of genomic DNA was performed. The guinea pig unique sequence was used as a probe. The guinea pig, as do the myomorph rodents, possesses one copy of the BC1 RNA gene (Figure 7).

Has the neural specific distribution shown for rat, mouse and hamster BC1 RNA (results not shown) also been conserved in guinea pigs? As shown in figure 8, the brain specific pattern is also maintained for this rodent. This indicates that despite its evolutionary distance and claims that the guinea pig is not a rodent (Graur et al., 1991), the BC1 RNA molecule has been transcribed in a brain-specific manner for at least 55 million years.

Figure 7. The guinea pig possesses a single copy of the BC1 RNA gene. Guinea pig genomic DNA was digested with a number of different restriction enzymes and electrophoresed on an agarose gel. The DNA was then hybridized with a probe complementary to the unique 3' end of the guinea pig BC1 RNA sequence and extending into the first 15 nucleotides of the 5' ID sequence. The presence of faintly hybridizing bands in addition to the main band suggests the presence of BC1 RNA pseudogenes.

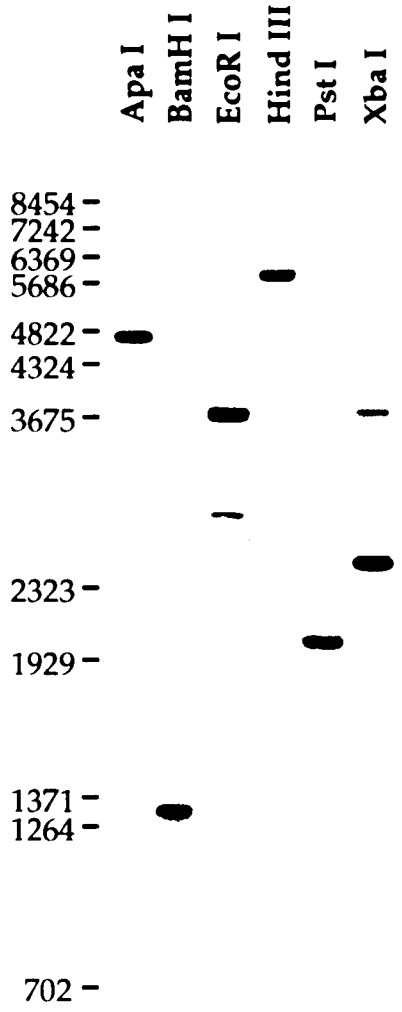
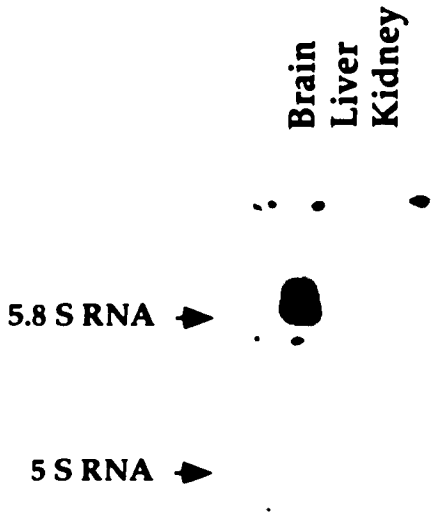

Apa I
BamH I
EcoR I
Hind III
Pst I
Xba I
8454
7242
6369
5686
4822
4324
3675
2323
1929
1371
1264
702

Figure 8. Guinea pig BC1 RNA is transcribed in a neural specific manner. Total RNA was isolated from three guinea pig tissue samples and subjected to electrophoresis on a denaturing acrylamide gel. The RNA was hybridized with a probe complementary to the unique 3' end of the guinea pig BC1 RNA sequence.


Brain
Liver
Kidney
5.8 S RNA
5 S RNA

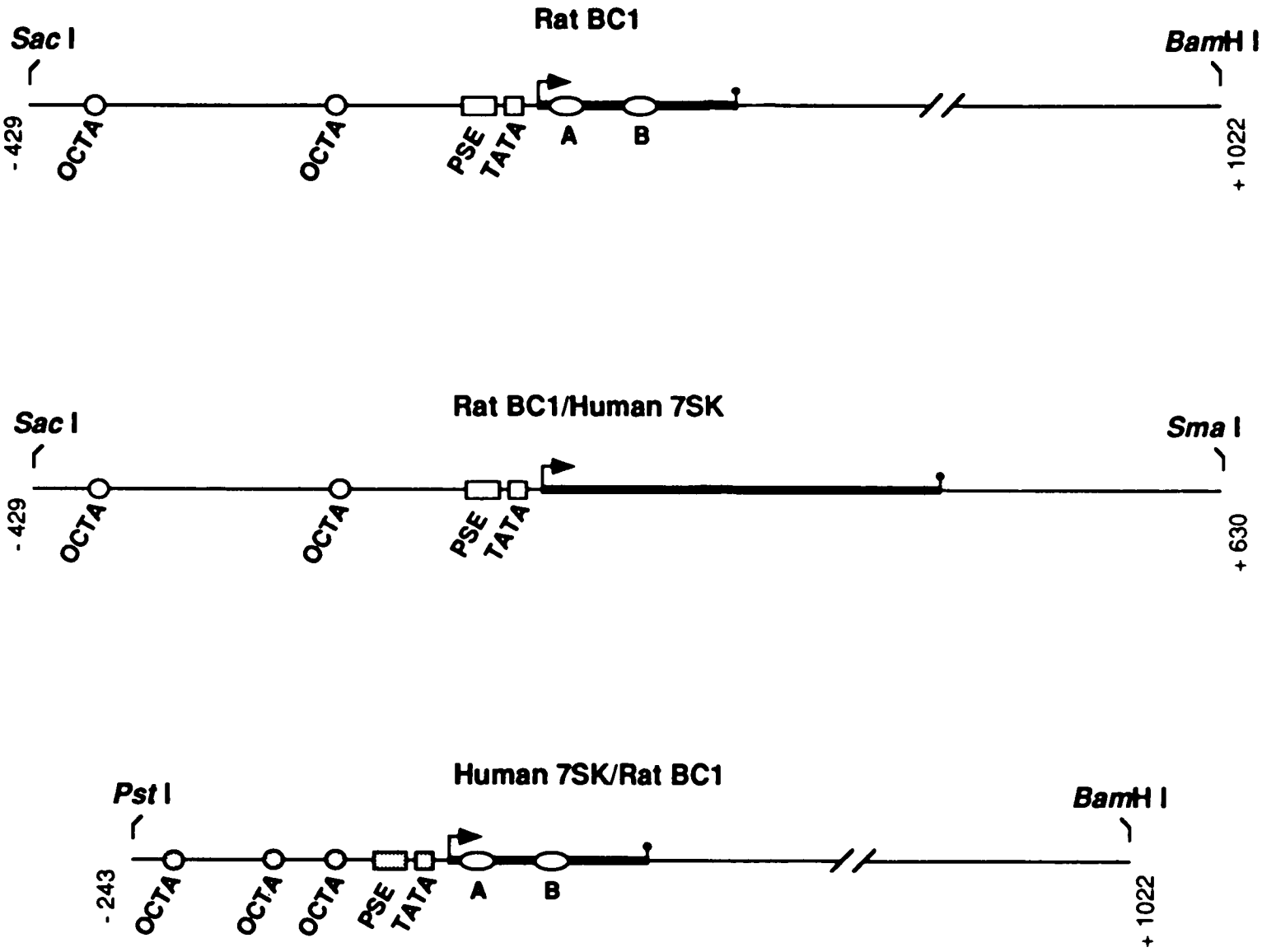
The establishment of rat BC1 RNA transgenic mice lines

One aim was to localize the *cis*- and *trans*-acting factors involved in both the developmental and neural-specific expression of BC1 RNA. By studying the *in vitro* differential transcriptional efficiencies of various BC1 constructs in whole cell brain extracts (Chapter 5), we hoped to identify sequences involved in determining transcriptional regulation. Once identified, constructs employing mutations in these regions or chimaeric constructs possessing BC1 RNA regulatory sequences could be directed towards transgenic research.

As a first step towards utilizing transgenic mice, three constructs were initially attempted for microinjection (Figure 9). The first was the rat BC1 gene with 429 bp of 5' flanking sequence and 1022 bp of 3' flanking sequence. The second and third genes were chimaeric constructions with rat BC1 and human 7SK (a plasmid containing the 7SK gene was kindly contributed by S. Murphy, Rockefeller University). The 7SK RNA, present in a variety of higher eukaryotes, is an abundant and stable nuclear RNA of unknown function which is transcribed by RNA pol III (Zieve and Penman, 1976; Reddy et al., 1981; Murphy et al., 1984). Transcription of human 7SK RNA *in vitro* and *in vivo* is solely dependent upon upstream promoters (Murphy et al., 1987; Kleinert and Benecke, 1988). These two constructs involved the swapping of the 5' flanking regions of these two genes with each other.

Purified DNA was prepared and given to the Mount Sinai Transgenic Facility for microinjection. Of an initial total of 90 mice born, from all three constructs, none were detected as transgenic.

Figure 9. Chimaeric constructs used for transgenic studies. The three constructs which were microinjected are shown drawn to scale. The human 7SK RNA gene was kindly provided by S. Murphy, Rockefeller Univ. (Murphy et al., 1987). The constructions are described in the text.


Rat BC1
Sac I
BamH I
- 429
OCTA
OCTA
PSE
TATA
A
B
+ 1022
Rat BC1/Human 7SK
Sac I
Sma I
- 429
OCTA
OCTA
PSE
TATA
+ 630
Human 7SK/Rat BC1
Pst I
BamH I
- 243
OCTA
OCTA
OCTA
PSE
TATA
A
B
+ 1022

However, from the last 14 samples analyzed, a total of 4 BC1 RNA wild type construct positive mice were identified (figures 10 and 11). These mice, two males and two females, were mated and each founder has produced a set of offspring. Analysis of these offspring, including Southern blots of DNA from each pup to determine the copy numbers present and Northern analysis of RNA extracted from different tissues to determine expression patterns, has not yet been performed. Lineages of each group of offspring, by intra-sibling crosses, will be generated.

Figure 10. Agarose gel electorphoresis of transgenic tail samples. The fourteen samples of DNA were digested with the restriction enzyme EcoR I and separated by electrophoresis. The gel was photographed after staining with ethidium bromide. Equivalent amounts of sample DNA and comparable levels of digestion are revealed.

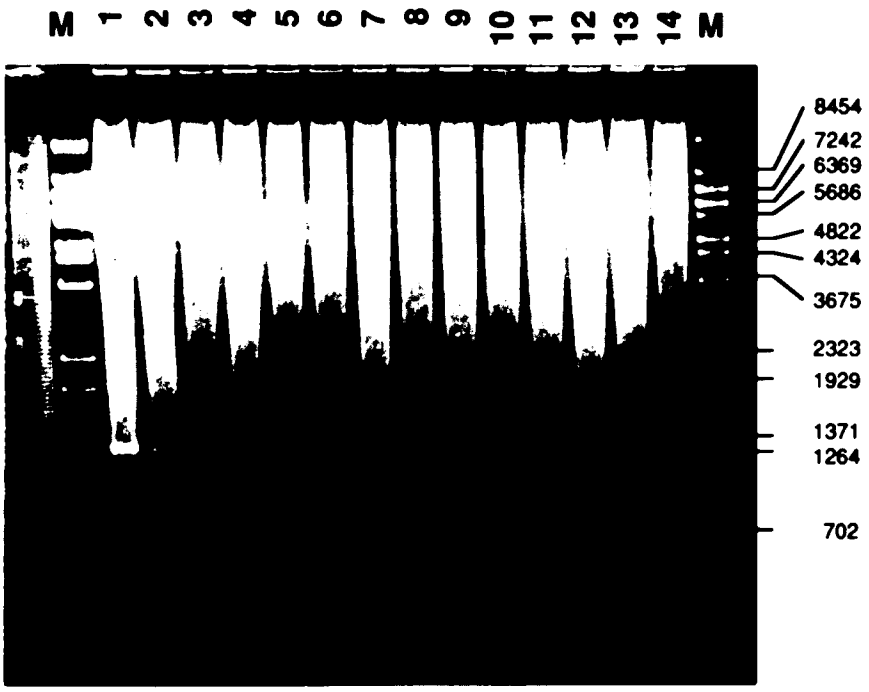
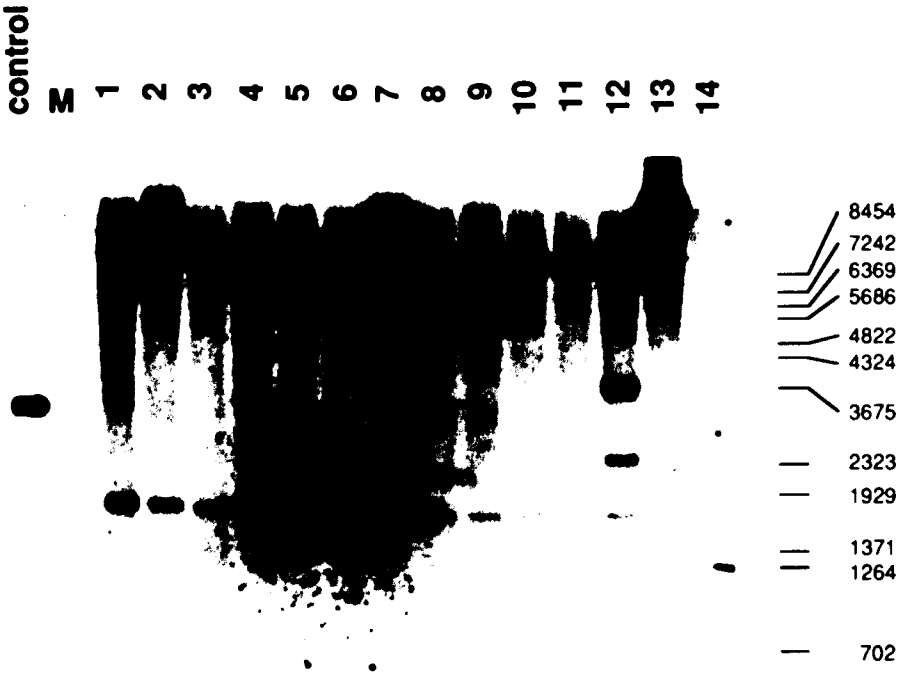

M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 M
8454
7242
6369
5686
4822
4324
3675
2323
1929
1371
1264
702

Figure 11. Southern analysis of transgenic DNA samples. The digested and size separated samples shown in figure 10 were transferred to a nylon membrane and hybridized with the upstream region (position -423 to -5) of the rat BC1 RNA gene sequence. The band present at 1700 bp represents the EcoR I fragment of the mouse BC1 RNA gene. The four transgenic mice, lanes 2, 7, 12, 13, are revealed by the presence of additional hybridizing bands.


control
M
1
2
3
4
5
6
7
8
9
10
11
12
13
14
8454
7242
6369
5686
4822
4324
3675
2323
1929
1371
1264
702

CHAPTER 4

EVOLUTIONARY IMPLICATIONS OF THE CONSERVATION OF THE BC1 RNA GENE

Guinea Pig remains a Rodent: Neural BC1 RNA as a Shared Marker

John A. Martignetti

and

Jürgen Brosius

Fishberg Research Center for Neurobiology

Box 1065

Mount Sinai School of Medicine

1 Gustave L. Levy Pl.

New York, NY 10029

Recent molecular phylogenetic analysis challenges the traditional placement of guinea pig-like rodents (caviamorpha) as a rodent suborder (Hystricognathi). The Hystricognathi would be provocatively placed into a separate order, possibly even more closely related to other mammals than those from their current suborder Sciurognathi, which includes the myomorphs, such as rats, mice, hamsters. We demonstrate here that BC1 RNA, a neural specific, small cytoplasmic RNA located in dendrites of a specific subset of neurons, is expressed in brains of both myomorph rodents and guinea pig, but not in lagomorphs, artiodactyls, or primates. The RNA sequence is conserved among the Hystricognathi and Sciurognathi at homology levels ranging between 90-99%. In addition, BC1 RNA's neural specific expression is maintained. These molecular findings support the morphological evidence for monophylie of Rodentia inclusive of guinea pig. This study also demonstrates the value of taxon-specific "marker molecules" in addition to "informative" mutations for complementing morphologically based relationships, especially when comparing species with fast rates of nucleotide substitutions.

The order Rodentia contains approximately half of all mammalian species (1) and based on a large body of morphological data has been subdivided into two suborders (2). The Sciurognathi, or Old World rodents, include the rat-like and the squirrel-like rodents (myomorphs and sciuromorphs, respectively). The Hystricognathi, or New World rodents, include guinea pig (*Cavia porcellus*) and chinchilla (*Chinchilla laniger*).

While these assignments are well grounded based on comparative morphology (3), the phylogenetic relationship based on the analysis of molecular information seemingly disturbs the older consensus. Namely, most available protein and nucleic acid sequences from guinea pig exhibit an unusual level of dissimilarity when compared to their myomorph homologues (4).

Recently, this "molecular paradox" has been addressed by Graur et al. (5) and "carried to a new extreme" (6). Using the maximum parsimony method they analysed published amino acid sequences and suggested that Hystricognathi diverged from the myomorph rodents before the separation of primates and artiodactyls. The authors therefore proposed that Hystricognathi should be elevated in taxonomical rank and be considered a separate and distinct mammalian order. This placement was both corroborated and disputed in a separate study by Allard et al. (7), who performed their own parsimony analysis based on published mitochondrial 12S ribosomal RNA gene sequences. Depending upon whether or not a distantly related avian sequence is used as an outgroup, either result, rodent polyphylie or monophylie, can be achieved. Both groups underscored the potential for error inherent to their method of analysis (5, 7) and both studies highlight the need for caution in inferring phylogenies from sequence analysis (see also ref. 8).

The observed high level of sequence divergence between the Sciurognathi and Hystricognathi suborders of rodents can be attributed to a fast rate of nucleotide substitution and as has been previously noted (5), unequal evolutionary rates can mislead parsimony inferences. "Time-landmarks of evolution" (9) or "genomic morphology markers" would therefore be useful tools to establish and confirm evolutionary relationships

in such instances. These may consist of repetitive sequences such as Alu elements (10), tRNA-derived retroposons (11), retropseudogenes or different expression patterns of genes in different species (12). We have investigated the value of BC1 RNA, a neural specific small cytoplasmic RNA (13) present in the somata and dendrites of a subset of neurons (14), as a taxon-specific marker molecule and applied it to the phylogeny of Sciurognathi and Hystricognathi.

BC1 RNA has been previously detected in myomorph rodents other than rats and its brain-specific expression pattern shown to be conserved (15). We examined whether BC1 RNA was also present in more distantly related mammals. An oligonucleotide probe complementary to the 3' terminal 60 nucleotides of rat BC1 RNA was used in RNA blots at very low stringency. No similarly sized RNA was detectable in samples from rabbit or bovine tissue (Fig. 1A). Only one hybridizing band was seen in human tissue. BC200 RNA (16) exhibits a cross-hybridizing signal consistent with the low degree of sequence similarity, restricted to its 3' end, with rat BC1 RNA (17; Figure 2). Intriguingly, BC200 RNA, a small cytoplasmic RNA, localized in the dendrites of an analogous subset of neurons similar to BC1 RNA (18), nonetheless has a distinct evolutionary origin from rat BC1 RNA. BC200 RNA is an actively transcribed monomeric Alu element (19).

We next probed RNA extracted from guinea pig. An RNA species of the appropriate size and present only in brain but not liver or kidney was identified (Fig. 1B). Therefore, BC1 RNA, as well as its neural-specific expression pattern, has been conserved in guinea pig but is absent in lagomorphs, artiodactyls and primates.

To determine the extent of BC1 RNA sequence conservation, cDNA from myomorph rodents, mouse (*Mus musculus*), syrian golden hamster

(*Microcetus aureus*), and guinea pig was prepared (20). The derived BC1 RNA sequences are shown and aligned in Fig. 2. Among the myomorph rodents BC1 RNA sequences exhibit 98-99% homology and between myomorph and guinea pig sequences 89-90% homology. This high level of RNA sequence conservation is in marked contrast to the lower levels present in flanking regions of the respective BC1 RNA genes (21). Comparison of blocks of 100 bp of 5' and 3' immediate flanking sequence among myomorphs shows greater than 86% homology. However, when compared to guinea pig and despite the presence of putative transcriptional regulatory elements, homology drops in these regions to much lower levels. For example, the 5' and 3' immediate flanking regions (60 bp, each) in guinea pig exhibit approximately 60% homology with the corresponding myomorph sequences (21). Regions beyond these flanks are even less conserved and require the introduction of numerous gaps to be aligned. These results clearly demonstrate that selective pressure has existed for at least 55 million years to maintain the BC1 RNA coding, and to various degrees, flanking regions.

The BC1 RNA gene was most likely generated via the retroposition of a $tRNA^{Ala}$. The 5' domain of BC1 RNA shares approximately 80% sequence identity with $tRNA^{Ala}$ (22) and the A-rich region located at the 3' end of the BC1 RNA\$tRNA^{Ala}$ homology is a structural hallmark of retroposition (23). In the new genomic environment, transcription of the former tRNA terminates further downstream at a stretch of four thymidine residues, an RNA polymerase III termination signal consensus sequence (24). The internalization of the A-rich region and the acquisition of a 3' unique domain accounts for the additional length of BC1 RNA in comparison to $tRNA^{Ala}$ (Fig. 2). In this regard it should also be noted that the 5' region of

BC1 RNA is also homologous to the ID sequence (13). The ID sequences are middle repetitive elements found dispersed in the rat genome in 10^5 copies while the mouse, hamster and guinea pig genomes contain 10^4, 10^3 and 10^2 copies, respectively (25). Sequence analysis reveals that they most likely arose via retroposition of $tRNA^{Ala}$ (26) or BC1 RNA (13).

Three different scenarios for the evolution of the BC1 RNA gene are shown in figure 3. These are shown in relation to the possible taxonomic positions of the myomorph and guinea pig lines and, we believe, strongly argue in favor of rodent monophylie. The first and perhaps simplest scenario has the BC1 RNA gene being generated after the myomorph and guinea pig lines jointly diverged from other mammals. Transcriptional activation could then occur at any point prior to the suborders themselves branching (Fig. 3A). This scheme is incompatible with guinea pig occupying a separate mammalian order. Second, the BC1 RNA gene could have been generated prior to the branching of the myomorph and guinea pig lines. Transcriptional activation would either need to occur prior to suborder branching, consistent with monophylie (Fig. 3B, tree I) or would have to have been activated twice independently, to be consistent with rodent polyphylie (tree II). Duplication of neural specific activation in two separate lines and the continued transcriptional silence in remaining orders seems highly unlikely. The third and most unlikely scenario has both retroposition and transcriptional activation occurring prior to the branching of the myomorphs and guinea pig. The BC1 RNA gene was then deleted or transcriptionally inactivated before the branching of artiodactyls and primates (Fig. 3C). If the BC1 RNA gene was activated prior to the branching of Sciurognathi and Hystricognathi and has since then been under selective pressure, it would be difficult to reconcile either the total

loss of the gene or its transcriptional activity in the orders that branched off thereafter.

Our data does not exclude the existence in other mammals of functional BC1 RNA molecules whose sequences have so highly diverged during the last 60-100 million years as to preclude their detection by cross-hybridization. This divergence would however seem highly unlikely given the relatively short evolutionary time scale and the high level of sequence conservation shown by other vertebrate small RNAs across different species.

The possibility also exists that additional RNA molecules, non-homologous to BC1 RNA, may have been recruited, from different ancestral genes, into similar functions. Primate BC200 RNA is potentially a functional, but as shown by its limited sequence similarity, not a phylogenetic homologue. The cooptation of BC1 and BC200 RNAs into a function must therefore have occurred independently. RNAs can, as is the case with proteins, continually yield new variants by gene duplication via recombination (27) or retroposition (28). These variants are then available for recruitment into novel tasks. The recruitment of BC1 and BC200 RNAs may have possibly enhanced, in their respective species, the efficiencies of proteins or RNP complexes already present in a wide range of organisms. Interestingly, only ribosomal RNA as polysomes (29), the messenger RNAs for MAP2 and a subunit of the Ca^{2+}/ calmodulin dependent protein kinase type II (30) are known to share the highly unusual dendritic localization of BC1 and BC200 RNAs. The neuronal function(s) of rodent BC1 RNA and primate BC200 RNA may be species specific given the necessary numerous and separate events needed to establish analogous RNAs in all other mammalian or vertebrate lineages.

REFERENCES

1. J.L. Hartenberger, in *Evolutionary Relationships among Rodents: A Multidisciplinary Analysis* W.P. Luckett and J.L. Hartenberger, eds (Plenum, New York) pp. 1-33 (1985).
2. T. Tullberg, *Nova Acta Reg. Soc. Scient. Uppsala* **18**, 1-514 (1899).
3. W.D. Luckett and J.L. Hartenberger, in *Evolutionary Relationships among Rodents: A Multidisciplinary Analysis* W.P. Luckett and J.L. Hartenberger, eds (Plenum, New York) pp. 685-712 (1985).
4. G. Sarkar, D.D. Koeberl, S.S. Sommer, *Genomics* **6**, 133-143; J.J. Beintema and Neuteboom, *J. Mol. Evol.* **19**, 145-152; S.J. Chan, V. Episkopou, S. Zeitlin, S.K. Karathanasis, A. McKrell, D.F. Steiner, A. Efstratiadis, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 5046-5050 (1984).
5. D. Graur, W.A. Hide, W.-H Li, *Nature* **352**, 649-652 (1991).
6. M.J. Novacek, *Nature* **356**, 121-125 (1992).
7. M.W. Allard, M.M. Miyamoto, R.L. Honeycutt, *Nature* **353**, 610-611 (1991).
8. J. Felsenstein, *Ann. Rev. Genet.* **22**, 521-565 (1988).
9. N. Okada, *Current Opinion in Genetics and Development* **1**, 498-504 (1991).
10. S.C. Ryan and A. Dugaiczyk, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 9360-9364 (1989); G. Del Pozzo and J. Guardiola, *Immunogenetics* 31, 229-232 (1990).
11. Y. Kido, M. Aono, T. Yamaki, K. Matsumoto, S. Murata, M. Saneyoshi, N. Okada, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 2326-2330 (1991); T.J. McConnell, W.S. Talbot, R.A. McIndoe, E.K. Wakeland, *Nature* **332**, 651-654 (1988); D. Tronik, M. Ekker, F. Rougeon, *Gene* **69**, 71-80 (1988).
12. S. Boyer, Montagutelli, X., Gomès, D., Simon-Chazottes, D., Guénet, J.-L., Dupouey, P., *Mol. Brain Res.* **10**, 159-166 (1991); J. Piatigorsky and G. Wistow, Science 252, 1078-1079 (1991); R.A. Cuthbertson, S.I. Tomarew, J. Piatigorsky, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 4004-4008 (1992).
13. T. M. DeChiara and J. Brosius, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 2624-2628 (1987).

14. H. Tiedge, R.T. Fremeau, Jr., P.H. Weinstock, O. Arancio, J. Brosius, *Proc. Natl. Acad. Sci. U.S.A* **88**, 2093-2097 (1991).
15. C. Sapienza and B. St-Jacques, *Nature* **319**, 418-420 (1986).
16. J. B. Watson and J. G. Sutcliffe, *Mol. Cell. Biol.* **7**, 3324-3327 (1987).
17. A signal is detected only in human material when using a probe designed complementary to the 3' 30 terminal nucleotides of BC200 RNA.
18. H. Tiedge, W. Chen, J. Brosius, submitted.
19. J.A. Martignetti and J. Brosius, submitted.
20. An oligo(dC) tail was added to the 3' ends of total RNA extracted from mouse, syrian (golden) hamster and guinea pig brain and cDNA synthesis and cloning into λZAP phage (Stratagene, La Jolla, CA) was performed as previously described (13). The respective libraries were screened with oligonucleotide HT005 (5'-A_3GGTTGTGTGTGCC AGTTACCTTGT$_9$GGTCT$_5$G TTAT$_4$GTCT$_5$-3') and/or probes corresponding to the 5' domain of rat BC1 RNA (13). Inserts from 15 guinea pig clones and from 20 mouse and hamster clones were sequenced using the dideoxy chain termination method [F. Sanger, S. Nicklen, A.R. Coulson, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977); F. Toneguzzo, S. Glynn, E. Levi, S. Mjolsness, A. Hayday, *BioTechniques* **6**, 460-469 (1988)]. Sequences were initially compared or aligned using the program BESTFIT or PILEUP in the GCG suite of programs (version 7.0) [J. Devereux, P. Haeberli, O. Smithies, *Nucleic Acids Res.* **12**, 387-395 (1984)], made available by the Department of Biomathematics, Mount Sinai School of Medicine, and in some instances, further aligned manually.
21. J.A. Martignetti, H. Tiedge, J. Brosius, *J. Cell. Biol. Suppl.* **15D**, 72 (1991).
22. T. Russo, F. Costanzo, A. Oliva, R. Ammendola, A. Duilio, F. Esposito, F. Cimino, *Eur. J. Biochem.* **158**, 437-442 (1986).
23. A. M. Weiner, P. L. Deininger, A. Efstratiadis, *Ann. Rev. Biochem* **55**, 631-661 (1986).
24. D. F. Bogenhagen, D. D. Brown Cell **24**, 261-270 (1981).
25. J. Rogers Int. Rev. Cytol. **93**, 187-279 (1985); P. L. Deininger in Mobile DNA D. E. Berg and M. M. Howe, eds., (Washington D.C.: American Society for Microbiology) pp 619-636 (1989).

26. G. R. Daniels and P. L. Deininger Nature **317**, 819-822 (1985).
27. S. Ohno, *Evolution by gene duplication* (Springer, New York, 1970).
28. J. Brosius, *Science* **251**, 753 (1991).
29. G.A. Banker and O. Steward, *Trends Neurosci.* **15**, 180-186 (1992).
30. C.C. Garner, R.P. Tucker, A. Matus, *Nature* **336**, 674-677 (1988); K.E. Burgin, M.N. Waxham, S. Rickling, S.A. Westgate, W.C. Mobley, P.T. Kelly, *J. Neurosci.* **10**, 1788-1798 (1990).
31. Isolation of RNA, its electrophoresis, blotting to nylon membrane, immobilization by UV-irradiation, and hybridization was carried out using standard procedures (32). An oligodeoxynucleotide, HT005 (18), complementary to the sixty 3' most nucleotides in rat BC1 RNA was 5' end-labelled using standard techniques (32). The filter was washed at 42°C in 5X SSC and exposed for six hours on Fuji RX X-ray film with intensifier screen. Guinea pig brain, liver, and kidney RNAs (30 μg each) were separated on a 6% acrylamide gel (acrylamide/bis-acrylamide ratio: 19+1) containing 7M urea and transferred to Gene Screen. An oligonucleotide, GU033, complementary to the 3' region of guinea pig BC1 RNA (5' AAAGGTTGTTTGTGTGCGCAGTTACCT TGTTTG 3') was 5' end labelled with ^{32}P-rATP and used as a probe for hybridization at 50°C in 20% formamide (32). The final wash was in 0.5X SSC at 55°C and the filter exposed overnight on Fuji RX X-ray film with intensifier screen.
32. J. Sambrook, E.F. Fritsch, T. Maniatis, *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1989).
33. Two bands located between 18S RNA and the position where BC1 RNA would migrate are detected in the lane corresponding to rabbit brain. They are also present in RNA extracted from rabbit liver (not shown) and probably are unspecific signals due to the low washing stringencies employed.

Figure 1. A. RNA blot of rat, rabbit, bovine and human brain RNAs screened at low stringency (31). A probe complementary to the sixty 3' most nucleotides of rat BC1 RNA (14) detects a small RNA only in the outer lanes corresponding to rat and human brain RNA. As discussed in the text, the human signal corresponds to BC200 RNA. Arrows indicate the positions of 18S and 28S ribosomal RNA. B. RNA blot of guinea pig brain, liver and kidney RNA (31). A signal corresponding in size to BC1 RNA is detected in guinea pig brain but not liver or kidney. Arrows indicate the positions of 5S (120 nucleotides) and 5.8S (160 nucleotides) ribosomal RNA. The Northern blot in A. was performed by W. Chen.

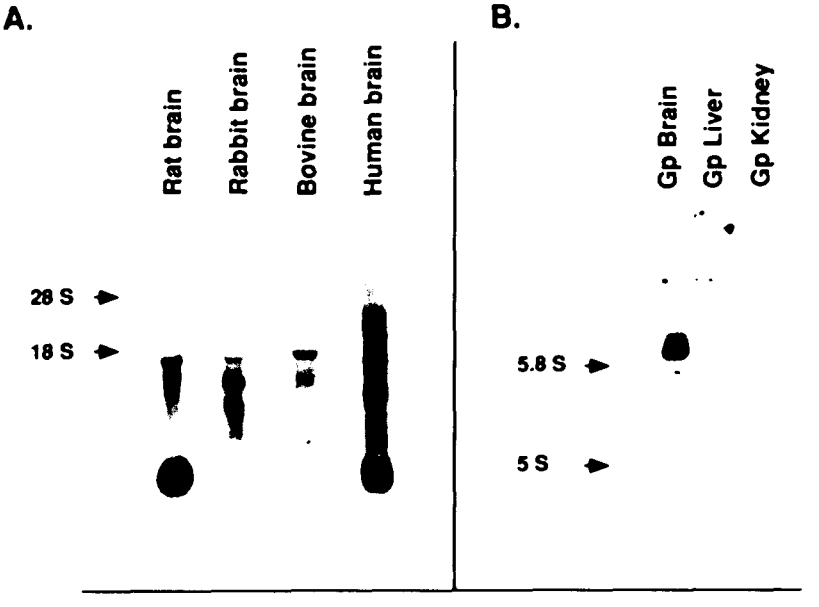

A.
Rat brain
Rabbit brain
Bovine brain
Human brain
28 S
18 S
B.
Gp Brain
Gp Liver
Gp Kidney
5.8 S
5 S

Figure 2. cDNA sequence alignment of BC1 RNA from rat (Rno), mouse (Mmu), syrian golden hamster (Mau) and guinea pig (Cpo). The sequences of rat BC1 RNA and mouse (Mmu) $tRNA^{Ala}$ are from references 13 and 19, respectively. The length of the poly(A) stretches have been derived from genomic sequences (21) except for syrian hamster which is indicated by A_n. The exact 5' end of rat BC1 RNA has been determined by RACE PCR (W. Chen, personal communication) while the 5' ends of the remaining cDNAs have been inferred from genomic sequences (21) and are indicated by "^". Numbering is with reference to the rat BC1 RNA sequence. The nucleotides that differ in mature mouse $tRNA^{Ala}$ (22) from the 5' domain of BC1 RNAs are shown. The human (Hsa) BC200 RNA region (pos. 134-178, ref. 18) that has 78% sequence identity to rat BC1 RNA is demarcated by slashes, "/". For all sequences, dots indicate identity and hyphens represent gaps introduced for optimal alignment.

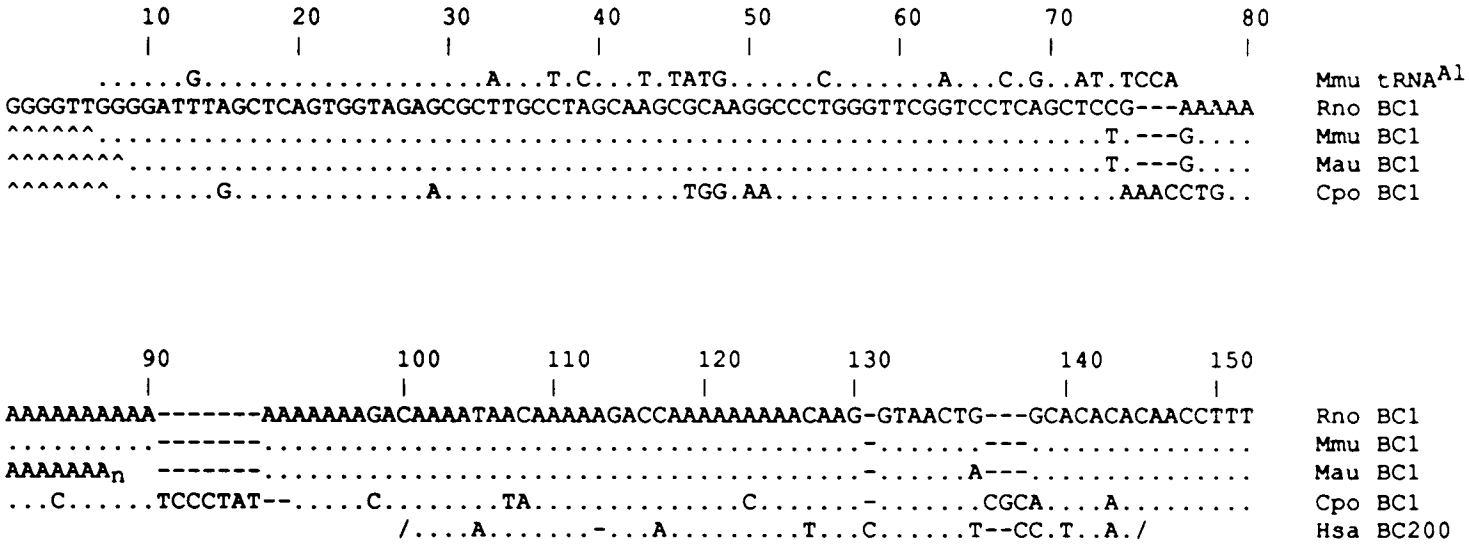
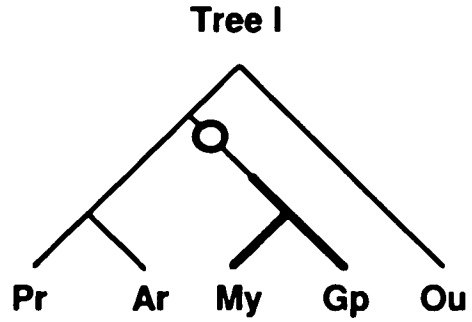
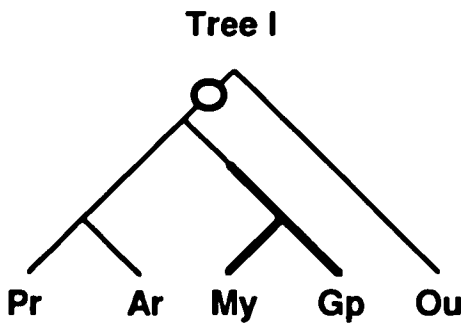

10 20 30 40 50 60 70 80
.....G.....A...T.C...T.TATG.....C.....A...C.G..AT.TCCA Mmu tRNA^Ala
GGGGTTGGGGATTTAGCTCAGTGGTAGAGCGCTTGCCTAGCAAGCGCAAGGCCCTGGGTTCGGTCCTCAGCTCCG---AAAAA Rno BC1
^^^^^^.....T.---G.... Mmu BC1
^^^^^^^^.....T.---G.... Mau BC1
^^^^^^^.....G.....A.....TGG.AA.....AAACCTG.. Cpo BC1
90 100 110 120 130 140 150
AAAAAAAAAA-------AAAAAAAGACAAAATAACAAAAAGACCAAAAAAAAAACAAG-GTAACTG---GCACACACAACCTTT Rno BC1
.....-------.....-.....---..... Mmu BC1
AAAAAAA_n -------.....-.....A---..... Mau BC1
...C.....TCCCTAT--.....C.....TA.....C.....-.....CGCA....A..... Cpo BC1
/....A.....-...A.....T...C.....T--CC.T..A./ Hsa BC200

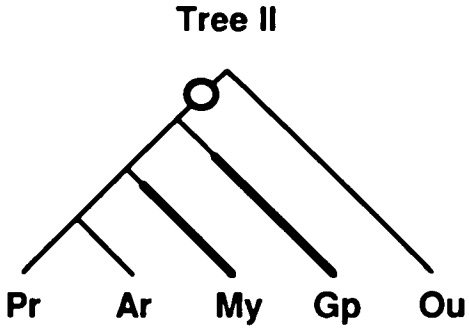
Figure 3. Phylogenetic trees for guinea pig (Gp), myomorph rodents (My), artiodactyls (Ar), primates (Pr) and an outgroup (Ou) (adapted from ref. 5). Tree I represents rodent monophylie and tree II polyphylie. Open circles represent the generation of the BC1 RNA gene and darkened lines transcriptional activation. The different scenarios are discussed in the text.

A

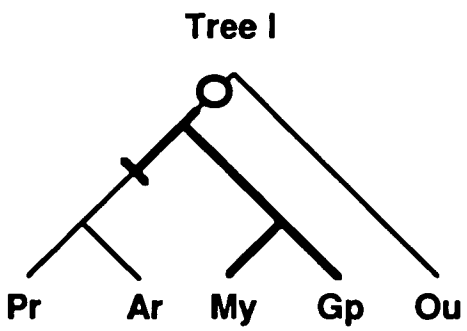

Tree I
Pr
Ar
My
Gp
Ou

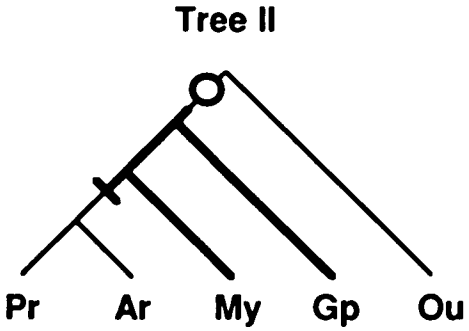
B


Tree I
Pr
Ar
My
Gp
Ou


Tree II
Pr
Ar
My
Gp
Ou

C


Tree I
Pr
Ar
My
Gp
Ou


Tree II
Pr
Ar
My
Gp
Ou

CHAPTER 5

The Neural Specific BC1 RNA Gene: An RNA Polymerase III Transcript with Functional Pol II and Pol III Promoter Elements

John A. Martignetti

and

Jürgen Brosius

Fishberg Research Center for Neurobiology

Box 1065

Mount Sinai School of Medicine

1 Gustave L. Levy Pl.

New York, NY 10029

SUMMARY

Rodent BC1 RNA represents the first example of a neural specific RNA polymerase III (pol III) transcription product. By developing a homologous *in vitro* system capable of supporting pol III directed transcription we are able to show that the BC1 RNA intragenic sequences as well as its upstream region, containing TATA, PSE and octamer binding sites, are both functional and necessary for transcription. Furthermore, the BC1 B box, differing from consensus B boxes of tRNAs by lacking an invariant A residue, represents a functionally related but distinct promoter element. The activity of the BC1 B box element is greatly increased, both in a BC1 RNA and a $tRNA^{Leu}$ gene construct, only when the BC1 5' flanking region is present and is located a specific distance from the intragenic A box. Moreover, a consensus B box sequence can only functionally replace the BC1 B box if the BC1 upstream region is removed. This "communication" between upstream and intragenic promoters results in either positive or negative interactions, suggesting a mechanism by which the BC1 RNA gene and possibly other pol III transcribed genes can be differentially regulated.

INTRODUCTION

Rodent BC1 RNA is a small, developmentally regulated, non-translated cytoplasmic RNA present in the somatic and dendritic compartments of a distinct subset of neurons in both the central and peripheral nervous systems (Sutcliffe et al., 1984; Tiedge et al., 1991). The RNA is transcribed from a single gene present in the genomes of rodents and is itself a founder, via retroposition, of the ID elements: a class of middle repetitive sequences present in 10^5 copies throughout the rat genome (DeChiara and Brosius, 1987; Martignetti et al., 1992). The BC1 RNA gene and the ID elements were themselves originally derived from a tRNA molecule (Daniels and Deininger, 1985; Sakamoto and Okada, 1985; Lawrence et al., 1985). It has been hypothesized that BC1 RNA, as part of a ribonucleoprotein complex, plays a role in either the pre- or posttranslational processes in neural postsynaptic compartments (Tiedge et al., 1991).

Neural BC1 RNA represents a unique model for transcriptional studies. It is a tissue specific, developmentally regulated RNA transcribed by RNA polymerase III (pol III). While RNA pol III synthesizes most small nuclear and cytoplasmic RNAs (reviewed by Ciliberto *et al.*, 1983 and Geiduschek and Tocchini-Valentini, 1988), only a few genes are transcribed by this polymerase in either a tissue or developmental specific manner. Examples include the oocyte specific and somatic 5S RNA genes of Xenopus (Brown, 1982), the $tRNA^{Ala}$ gene of *Bombyx mori* (Sprague et al., 1977; Meza et al., 1977) and the primate neural specific BC200 RNA (Watson and Sutcliffe, 1987; Martignetti and Brosius, 1992), where regulation most likely occurs at the level of transcription. The promoters of pol III transcribed genes (class III) which are both necessary and sufficient for transcription, were

originally found to be highly conserved intragenic sequences (Sakonju *et al.*, 1980; Bogenhagen *et al.*, 1980; Kressman *et al.*, 1979; De Franco *et al.*, 1980; Galli *et al.*, 1981; Hofstetter *et al.*, 1981; Sharp *et al.*, 1981; Fowlkes and Shenk, 1980). These intragenic promoters are present in the majority of class III genes and two noncontiguous domains, the A and B boxes, have been defined for tRNA, adenovirus VA RNA genes, the Epstein-Barr EBER RNA genes as well the human and rodent Alu repetitive sequences (Ciliberto *et al.*, 1983; Geiduschek and Tocchini-Valentini, 1988). The 5S RNA genes in addition to possessing an A box contain a 5S unique sequence, the C box (Geiduschek and Tocchini-Valentini, 1988). However, recent work has made it evident that: 1. numerous pol III genes require in addition to intragenic promoters upstream sequences, which quite surprisingly, are similar not only in sequence but also in position to well characterized pol II promoters; and, 2. some pol III genes require only the sequences residing upstream from the coding sequence (reviewed in Sollner-Webb, 1988 Folk, 1988, Murphy *et al.*, 1989, and Kunkel, 1991).

We have recently identified and sequenced the BC1 genes from several rodent species including rat, mouse, Chinese and Syrian hamsters and guinea pig (Martignetti et al., 1992). For at least 55 million years, pressure has been maintained to conserve the rodent BC1 RNA coding and flanking sequences and their neural-specific expression pattern. Sequences conserved within the coding region of the BC1 gene include regions resembling the A and B boxes of tRNAs. Unexpectedly, the conserved upstream flanking regions of these genes possess several sequences resembling RNA polymerase II (pol II) promoter sequences: including two octamer binding sequences, a proximal sequence element

(PSE), and a TATA box. To evaluate the possible involvement of these sequences in regulating BC1 RNA transcription we have undertaken the present study.

In this paper we report the development of a homologous whole cell *in vitro* extract, prepared from rat brain tissue, capable of supporting pol III transcription. The rationale was to eventually study the tissue specific transcriptional regulation of the BC1 RNA gene while avoiding the use of cultured rodent cell lines, which curiously exhibit deregulated BC1 RNA expression patterns (McKinnon *et al.*, 1986). Using extracts prepared from rat cortex (a brain region expressing BC1 RNA), in contrast to those obtained from HeLa cells (a heterologous source used in many transcription assays), we demonstrate that the BC1 RNA gene absolutely requires not only its pol III intragenic sequences but also upstream sequences for transcription. Removal of the entire upstream region abolishes transcription while mutation of the TATA and PSE sequences greatly diminishes transcription efficiency. Furthermore we are able to show that the BC1 B box represents a variant intragenic promoter. This element requires the presence of the BC1 upstream region to amplify transcription. Replacement of the BC1 B box with a consensus B box sequence results in a severe loss of transcriptional efficiency which can be restored if the BC1 upstream region is removed. These results may be explained by the presence of positive and negative interactions between competing factors, suggesting that communication between the intragenic and upstream promoter regions of the BC1 RNA gene can regulate promoter activity and gene expression.

RESULTS

Preparation of an *in vitro* transcription extract and optimization of transcription parameters.

A series of whole cell extracts were prepared from rat brain cortex as described in the Experimental Procedures section. These extracts have allowed the study of transcriptional regulation of a neural-specific class III gene, whose transcription is found to be deregulated in cultured cells. The methodology is based upon the whole cell extract of tissue culture cells pioneered by Manley for RNA polymerase II studies (1980, 1984); itself a modification of the method of Sugden and Keller (1973) for isolating RNA polymerase II. Modifications have been introduced in both extract preparation and assay conditions for reproducibility with samples of neural origin. Application of this method to liver has yielded inconsistent results. One possibility for this inconsistency may be an excess of nucleases present in liver relative to brain tissue extracts. Exogenously added template DNA was almost totally degraded with liver but not brain extract at 30° C and only partially degraded with either extract when incubated on ice (results not shown). The addition of EDTA to the liver extract (5 mM final concentration) resulted in inhibiting template degradation but shifted the template to a predominant form with much slower mobility than that of the control template as viewed on agarose gels.

Optimization of assay conditions for brain extracts were first determined. For most templates, increasing the amount of extract increased the amount of transcription product formed. The amounts tested varied from 20% to 75% of the final transcription reaction volume.

A volume of half the final volume was used as this amount of extract was in equilibrium with the amount of product formed while allowing the final $MgCl_2$ concentration to be kept to a minimum. Altering $MgCl_2$ concentrations yielded variable results in both product amount and for the tRNA templates, size of transcript generated (assumed to be a result of variability in start site selection; Figure 1A). Optimum values for templates tested varied between 1 mM and 4 mM. The final KCl reaction concentration was varied from 50 mM to 100 mM with no changes observed in the amount of product formed.

The effective range of template concentration used varied between 10 and 30 μg/ml. Adding increasing amounts of template DNA increases and then decreases the amount of specific product suggesting that proteins necessary for specific transcription are being titrated by the excess template. This value varied by preparation and by template tested. Temperature (Figure 1B) and pH parameters were optimized and are similar irrespective of the template tested.

We compared the ability of the whole cell cortex extracts versus extracts prepared from HeLa cells (Dignam et al., 1983) to initiate specific transcripts on exogenously added templates containing various pol III RNA genes: Xenopus t^{Met} (Clarkson et al., 1978), rat t^{Leu} (Makowski et al., 1983), human 7SL (Ullu and Weiner, 1985) and rat BC1. Each template directs the synthesis of a product which is indistinguishable, except for the total amount synthesized, from either extract. At present we are unable to explain why several different cortex extracts were unable to support transcription from either a Xenopus 5S RNA or human 7SK RNA gene containing template. Finally a comparison between these whole cell extracts and extracts prepared

from isolated rat cortex nuclei (Gorski et al., 1986) was made on a panel of different BC1 mutant templates (BC1:KS, gBC11, ΔTATA, and ΔPSE 1 and 2; described in the text) and a $tRNA^{Met}$ gene. No differences were seen in the transcriptional efficiencies of the various templates between prepared extracts (results not shown) implying that the whole cell extracts are as reliable as nuclear derived extracts for trancriptional analysis of different BC1 RNA and tRNA gene constructions.

The BC1 sequences resembling intragenic box A and B are necessary for transcription.

The BC1 RNA gene and its upstream sequence is shown in figure 2. Within the RNA coding sequence are regions resembling the internal control regions of tRNAs, the A and B boxes. This is most likely the result of the BC1 gene's evolution via a tRNA (Martignetti et al., 1992). For tRNAs these promoters have been defined as necessary for transcription (Galli *et al.*, 1981; Hofstetter *et al.*, 1981; and Sharp *et al.*, 1981). We decided to test whether the BC1 RNA sequences have maintained their transcriptional function and affect the *in vitro* levels of BC1 RNA synthesis (Figure 3).

For the first set of constructs examined, the A and B box sequences were deleted independently (constructs BC1: -(A), -(B)) and both together (construct BC1:-(A+B)). These constructs were tested for their ability to direct RNA synthesis in the cortex and HeLa derived extracts. In the cortex extract, the loss of the A box (Figure 3A, lane 2) leads to transcriptional inactivity whereas in the HeLa extract, this same deletion only yields a slight decrease in transcriptional efficiency (Figure 3C, lane 2). However for both extracts, BC1 RNA synthesis is no

longer detectable with the loss of the B box or with the loss of both A and B boxes (lanes 3 and 4 of Figures 3A and 3C, respectively).

Because these deletion constructs produce shortened transcripts, it may be argued that this could result in an instability of the synthesized RNA or of the DNA template, therefore two other constructs were created. In this second set, mutations were created which did not delete but instead inserted a novel sequence into the A and B box regions (constructs BC1: Δ A and Δ B). Nonetheless, as shown in the second and third lanes of figures 3B (cortex extract) and 3D (HeLa extract), these constructs gave similar results as to the original deletion constructs. Therefore, and regardless of which extract is used, the tRNA homologous internal control regions of the BC1 RNA gene are necessary for transcription and are therefore functionally homologous to those for other tRNA genes.

In most class III genes the B box is the primary determinant of promoter strength while the A box, which is dispensable, determines the initiation site (reviewed in Geiduschek and Tocchini-Valentini, 1988). Under optimal assay conditions the A box can also determine promoter strength (Geiduschek and Tocchini-Valentini, 1988). An example is provided by the effect of box A deletion and mutation on the synthesis of EBER2 RNA (Howe and Shu, 1989). When analyzed *in vivo* no transcription was obtained from a construct which lacked the A box, whereas *in vitro*, a mutated A box construct gave overall levels of synthesis comparable to wild type expression. Therefore, the differences seen between extracts on the deleted/mutated BC1 A box templates may be indicative of the greater fidelity of the tissue prepared extracts.

The BC1 B box represents a variant internal control element

By comparing the A and B boxes from numerous tRNA genes a number of 'invariant' residues have been identified within these intragenic promoters (reviewed in Sharp et al., 1984). This conservation possibly reflects not only the necessity for certain nucleotides to promote transcription but also for these sequences to form part of the RNA product, which must be able to form higher-order structures and also interact with other components of the protein synthesizing machinery. While the BC1 B box is strikingly similar to the consensus derived for tRNA B boxes, direct comparison reveals that an invariant A residue is lacking from the BC1 sequence (Table I). For the $tRNA^{Pro}$ gene, A→G transitions or A→T transversions from the invariant A residue can reduce transcription rates two-fold (Traboni et al., 1984). Similar decreases are seen for A→T transversions using the Alu intragenic promoter (Murphy and Baralle, 1983). Consequently, we wondered what effect "correction" of the BC1 B box sequence to that of the consensus, by inserting the invariant A residue, would have on the *in vitro* BC1 RNA transcription levels.

If the BC1 B box sequence is mutated to match the consensus sequence, expression levels are not increased but paradoxically the levels are drastically reduced (Figure 4, lane 3). Yet this level is nonetheless higher than that achieved with either the mutated (lane 2) or deleted (lane 4) B box constructions. Therefore, while the BC1 B box, or b' box, is functionally similar to the consensus B box of tRNAs in that it is required for transcription these two promoter elements are not mutually interconvertable and could represent different binding sites.

The BC1 b' box and a correctly positioned BC1 5' flanking region can combine to increase transcriptional activity

Why does the mutation of the BC1 b' box to the B consensus sequence result in the decreased, but not loss of, synthesis of BC1 RNA? The 5' regions for some 5S genes (Morton and Sprague 1984; Selker *et al.*, 1986; Garcia *et al.*,1987), tRNA genes (Sprague *et al.*, 1980; Dingermann *et al.*, 1982; Hipskind and Clarkson, 1983; Johnson and Raymond, 1984), the 7SL (Ullu and Wiener, 1985) and EBER I and II genes (Howe and Shu, 1989) have been shown to play a modulatory role in transcriptional activity (reviewed in Kunkel, 1991). We wondered whether the BC1 5' region is also somehow involved in transcriptional regulation and thereby "dissonant" with the presence of a consensus B box resulting in a block to transcription.

To assess this, the B box consensus sequence was placed into a BC1 gene construct which lacked any 5' flanking sequence (BC1:con B: del 0). The result as shown in figure 5A (lane 3), is that a B box consensus sequence is functional when the BC1 upstream flanking sequence is removed.

This result suggested that the BC1 upstream and intragenic control regions form a distinct complex. Can the spacing between these elements be important to their function? A series of constructs with a rat $tRNA^{Leu}$ gene were prepared to examine this possibility.

To establish a baseline for the $tRNA^{Leu}$ transcription levels, two initial constructs were made and assayed as shown in Figure 5B. The tRNALeu gene yields relatively low transcription levels (compare lanes 1 and 2). First, the effect of mutating the rat $tRNA^{Leu}$'s B box from a consensus sequence to one more closely resemble that of BC1 RNA (construct tRNA b') was examined. In the cortex extract, deletion of the

invariant A residue did not affect transcription (compare lanes 2 and 3). Second, the BC1 upstream sequence was exchanged for that of the tRNA's (construct BC1/tRNA). This construct (lane 4), possessing the tRNA's consensus B box sequence, directs the synthesis of an equivalent amount of correctly terminated product, however, the total amount of radiation incorporated is decreased as the higher molecular weight products seen in the adjacent two lanes is not present. These longer transcripts most likely represent incorrectly terminated product. Therefore the fusion of the BC1 upstream region to the $tRNA^{Leu}$ gene results in an overall decrease in transcription.

Finally, the BC1 5' region was placed upstream of the tRNA gene containing the BC1 b' box. The first construction, BC1/tRNA b', fused the 5' flank onto the tRNA with the BC1 TATA box 5 nt closer to the tRNA A box than in the BC1 gene. The second, BC1 +5/tRNA b', placed the TATA box at an equivalent position to that of the BC1 RNA gene. As shown in the short time exposure in figure 5C, when the BC1 upstream region and the tRNAs A box are spaced as in the BC1 RNA gene, the result is a greater than 50-fold increase in transcription levels from wild type controls.

Upstream sequences are essential for the transcription of the BC1 gene

The results of these initial experiments reveals that not only are the BC1 intragenic control sequences necessary for transcription but also that the BC1 flanking region is involved in transcriptional regulation of BC1 RNA. Sequence analysis of the 5' flanking region of the BC1 RNA gene reveals the presence of several pol II promoter motifs

(Martignetti et al., 1992). These include two octamer binding sequences (positions -387, -178), a proximal sequence element (PSE, position -60), and a TATA box (position -28). Interestingly, these elements are also found in the upstream sequence of the pol III transcribed 7SK and U6 genes, whose promoters are located exclusively in their 5' flanking regions (Murphy *et al.*, 1987; Kleinert and Benecke, 1988; Carbon *et al.*, 1987; Das *et al.*, 1988; and, Kunkel and Pederson, 1988). To determine what role the upstream region of the BC1 gene plays in transcription, a series of progressive 5'-end deletion mutants were constructed (Figure 6A). These were then assayed, as shown in figure 6, for their ability to direct *in vitro* transcription.

When assayed in the cortex extracts (Figure 6B) wild type expression is maintained when the first 53 nucleotides are present. Deletions past nucleotide 53, deletions -33 (~ 50% of control) and -17 (< 5% of control), decrease transcription levels until finally transcription is virtually undetectable at deletion 0. Curiously, including 72 or more nucleotides raises trancription levels by 30%, but inclusion of more than 173 nucleotides, which contains an octamer binding sequence, decreases levels back to wild type control.

When assayed in the HeLa extracts, no effect is seen on the relative transcription levels from 5' deleted BC1 RNA gene constructs (Figure 6B). This difference in transcriptional requirements for the BC1 RNA gene depending on source of extract is similar to the results seen for the tissue specific *Bombyx mori* $tRNA^{Ala}$ gene (Sprague et al., 1980). For the BC1 RNA gene, as well as the $tRNA^{Ala}$ gene, only the homologous tissue prepared extract reveals its requirement for flanking sequence. These results stress the importance of homologous extracts

for examining regulatory and not basal specific transcription regulation and that the BC1 RNA gene requires upstream flanking information for transcription.

Defining the role of Pol II motifs within the upstream sequence

Within the BC1 upstream region defined as providing maximal transcription are located the TATA box and PSE elements. To better characterize what role the BC1 TATA and PSE sequences perform, mutations were created in these sequences and their effects on the levels of BC1 RNA synthesized were examined (Figure 7).

In the HeLa extracts (Figure 7B), mutation of both of these elements results in no loss of transcriptional efficiency unlike the results seen in the cortex extracts. In the cortex extract (Figure 7A), mutation of the TATA box from its wild type sequence significantly decreases the amount of BC1 RNA synthesized, but not to the levels observed when all upstream information is removed.

Efficient transcription is still achieved when the first 53 nt (Figure 6B, lane 5) of upstream information are present. This deletion occurs approximately at the midpoint of the PSE element; dissecting the element into a left and right half (Table II). Therefore it was quite unexpected that mutation of the left half of the PSE sequence, BC1:ΔL, a region missing from the -53 deletion construct, resulted in transcription levels much lower than wild type. Mutation of the right half of the PSE sequence, BC1:ΔR, results in slightly increased transcriptional efficiency. Based on the apparently contradictory results obtained from the -53 deletion and BC1:ΔL and ΔR mutations, it seems likely that a level of cooperativity is achieved between the different halves of the BC1 PSE

element and regions further upstream. A similar type of result was seen in experiments on the 7SK RNA gene (Murphy et al., 1989), however for this gene, mutation of the right half of the PSE decreased transcription levels more than mutations in the left half. Cooperativity between the right half of the PSE and the 7SK upstream region was revealed by stimulated transcription levels by exogenously added octamer transcription factor.

Although it contains similar pol II motifs, the BC1 upstream region cannot functionally replace that of the 7SK gene

As already stated, several of the 7SK upstream promoter motifs are also present in the BC1 gene: a TATA-like sequence, a snRNA proximal sequence element (PSE) and two octamer motifs. As further demonstrated, the PSE elements of the BC1 RNA and 7SK genes seem to share a functional interelationship with their upstream regions. The upstream region of the 7SK gene has been shown to direct transcription even in the complete absence of the 7SK RNA coding region (Murphy *et al.*, 1987; Kleinert and Benecke, 1988). To evaluate the extent of functional similarity between the upstream regions of the BC1 RNA and 7SK RNA genes, chimaeric constructs, swapping the upstream regions of these genes, were made and then assayed in HeLa extracts (results not shown). The BC1 RNA gene upstream region is unable to support transcription of the 7SK gene. Therefore, while structurally similar, these upstream regions are not functionally equivalent.

DISCUSSION

In this report we describe the preparation of a transcriptionally active extract prepared from rat brain cortex used to analyze the transcriptional regulation of the rat BC1 gene, the first identified neural specific pol III transcribed gene. To provide a means for investigating the mechanisms underlying BC1 transcription an easily prepared whole cell tissue *in vitro* system was defined that supports faithful pol III transcription. A whole cell extract was chosen because it has been shown in the past (Schwartz et al., 1974; Wu, 1978; Dignam et al., 1983) that appreciable amounts of RNA polymerase III and required factors are lost from the nucleus during its isolation.

The necessity for using tissue and not cultured cells as starting material for preparing transcriptionally active extracts is underscored by the differential transcription efficiencies seen between templates assayed in the whole cell extracts and extracts prepared from HeLa cells (see Figures 3, 6, and 7). Whereas the extracts derived from tissue allowed us to inspect regulatory sequences involved in BC1 RNA synthesis, the extracts prepared from cultured cells were less stringent in their requirements for the upstream BC1 promoter elements and therefore could not identify the interactions between intragenic and extragenic promoters. These *in vitro* differences seem to parallel the deregulation of BC1 RNA transcription seen in cultured rodent cell lines (McKinnon et al., 1986). Cultured cell types commonly used for extract preparation are generally poorly differentiated and only a few reports have suggested that tissue specificity might be reproduced under *in vitro* conditions (Tsuda and Suzuki, 1981; Bazett-Jones et al., 1985; Scholer and Gruss, 1985), as has been previously noted (Gorski et al., 1986).

Using the whole cell extracts we demonstrate that the BC1 RNA gene represents a class III gene in which the essential sequences necessary for transcription are not only intragenic but also reside in the 5' flanking region of the gene. The BC1 RNA gene intragenic sequences, while structurally similar to the A and B boxes of tRNAs, Alu sequences, and EBER and VAI genes, contains a variant promoter element such that the fully functional BC1 b' box and a consensus B box are not functionally interconvertable (see Figures 4 and 5). Perhaps the most intriguing feature of the b' box is that is only maximally functional in the presence of the BC1 5' flanking region and is critically dependent upon the spacing from it. A consensus B box is barely functional with the BC1 5' flanking region present while removal of the BC1 5' flanking region activates (de-represses) the consensus B box. These results strongly suggest that a juxtaposition of transcription factors, dependent on the combination of intragenic and extragenic sequences present and mediating either positive or negative interactions, are competing to form a transcription complex on the BC1 RNA gene template. A model encompassing these results is presented in Figure 8.

In addition to pol III, at least two common protein factors are required for the transcription of all class III genes, TFIIIB and TFIIIC (Segall et al., 1980). The primary step in initiation complex assembly in tRNA-like genes (Figure 8A) is the binding by TFIIIC to the internal control region, where the B box is the major determinant of binding affinity and therefore of promoter strength (Lassar et al., 1983). Once bound, TFIIIC cooperatively interacts with TFIIIB, which is itself unable to bind to DNA directly, to correctly position it upstream of the transcription start site (Lassar et al., 1983; Bieker et al., 1985). In yeast,

once TFIIIB has formed a stable complex it is resistant to removal by high salt concentrations and heparin, conditions which dissociate TFIIIA and TFIIIC, and is able to direct multiple rounds of transcription (Kassavetis et al., 1989 and 1990).

Contained within the BC1 upstream sequence shown to be necessary for transcription are regions which resemble, in sequence and organization, promoters for pol II dependent genes, including a TATA, PSE and OCTA elements (Figure 2). These pol II promoter elements have also been found in the upstream regions of the 7SK and U6 RNA genes (Murphy *et al.*, 1987; Kleinert and Benecke, 1988; Carbon *et al.*, 1987; Das *et al.*, 1988; and, Kunkel and Pederson, 1988). Therefore it is quite interesting that a negative transcriptional interaction between upstream and intragenic *cis*-acting elements in U6 chimaeric constructs has been reported by Parry and Mattaj (1990). When the upstream U6 sequence, containing the TATA and PSE elements, was fused to a tRNA gene or a construct containing functional intragenic box A and B sequences the template was transcriptionally inactive. Removal of any one of the four promoter elements or an increased spacing between them seemed to relieve the inhibition: indicating a competition between different assembling initation complexes. This result is echoed in the result obtained when a consensus B box was used to replace the BC1 b' box (Figure 8B). In the presence of the BC1 upstream region the B box was only barely functional. Removal of the upstream region relieved the inhibition and transcription levels increased dramatically. Another example by which a pol III transcribed gene can be negatively affected via steric hindrance has been described for a yeast $tRNA^{Glu}$ suppressor gene (Dingermann et al., 1992). In this experiment, the bacterial *tet*

operator was positioned upstream of the yeast tRNA start site. When the bacterial repressor protein binds to this operator, the level of $tRNA^{Glu}$ is decreased at least 50 fold.

The BC1 b' box is able to mediate, unlike the consensus B box, a positive interaction between an intragenic promoter and gene external promoters. A modified form of TFIIIC, which allows for both the binding to a related but distinct B box sequence and a sterically allowed positive interaction with an upstream assembled complex may be responsible for this effect (Figure 8C). In the absence of the necessary upstream sequences, cooperative or stabilizing interactions for the modified TFIIIC would be lost and binding would not be favored (Figure 8D). As shown for the tRNA construction, the spacing of the elements is also critical for transcriptional activation, such that altering the periodicity by half a helical turn (5 nucleotides) can inactivate promoter interaction (Figure 5B and C).

A related example of positive interaction between pol III intragenic and pol II extragenic promoters is provided by the EBER2 gene of Epstein-Barr virus (Howe and Shu, 1989). It contains not only functional A and B boxes but also functional upstream pol II promoter elements, a TATA box and Sp1 and ATF binding sites, which have been shown by DNAse I footprinting to be involved in protein binding. Mutations in these elements or deletion of the upstream region causes a loss of transcriptional efficiency when assayed *in vivo*, however, *in vitro*, no such losses are observed. What may be of particular relevance in comparison of this system to the BC1 RNA gene is that the EBER2 RNA gene's functional B box lacks certain invariant nucleotides (Table II).

Can modified forms of TFIIIC be responsible for both the binding to variant B boxes and communication with upstream elements? Developmental changes in TFIIIC have recently been demonstrated to be responsible for the differential expression of certain class III genes in *Xenopus* (Reynolds and Johnson, 1992). TFIIIC isolated from immature oocyte extracts but not from mature extracts activates two oocyte specific transcripts, $tRNA^{Met1}$ and OAX, whereas activation of the somatic 5S gene is achieved regardless of extract source. Seemingly, during *Xenopus laevis* development, changes occur in the activity of TFIIIC, by an unknown mechanism, which allows for the selective transcription of pol III transcribed genes. The activity of TFIIIC can also be regulated in response to viral products. This as shown by its increased activity in the presence of the E1A gene product of adenovirus (Hoeffler and Roeder, 1985 and Yoshinaga et al., 1986 and Hoeffler et al., 1988) and the X-gene product of the hepatitis B virus (Aufiero and Schneider, 1990) or in response to SV40 transformation (White et al., 1990) and its decreased activity in poliovirus infected cells (Fradkin et al., 1987). The activity of TFIIIC can also be modified *in vitro*: stimulation by dithiothreitol and inhibition by alkylating agents (Cromlish and Roeder, 1989). The structure of mammalian TFIIIC, unlike that for yeast (Johnson and Wilson, 1989; Gabrielson et al., 1989; and, Parsons and Weil, 1990) has remained elusive. However, common to all the different purification methods employed, multiple proteins, as with yeast, are present in the active fractions (reviewed in Palmer and Folk, 1990). Different combinations of these proteins, those responsible for DNA binding or those mediating protein-protein interactions, or the modification of

proteins in the complex could be methods by which variant forms of TFIIIC are generated.

Not only do the genes transcribed by pol III use many of the same strategies for transcription initiation complex formation as do the genes transcribed by pol II, but pol III transcribed genes may also use the same pol II components (reviewed in Murphy et al., 1989 and Gabrielsen and Sentenac, 1991). Several groups have now shown that the TATA-binding protein is involved in the *in vitro* transcription of class III genes (Lobo et al., 1991; Margottin et al., 1991; and Simmen et al., 1991). Another example of the interconnectedness of pol II and III transcription has been described wherein the polymerase specificity of several U RNA genes has been switched (Mattaj *et al.*, 1988; Lobo and Hernandez, 1989; Waibel and Filipowicz, 1990). It has now also been shown that the TATA-binding protein is also an integral component of an RNA pol I transcription factor (Comai et al., 1992). Therefore, the underlying "network of interactions" (Sharp, 1992) between the different polymerases becomes more apparent with the analysis and dissection of the transcription of different genes in defined biochemical systems.

Experimental Procedures

Construction of 5' and 3' deletion mutants.

A limited digest of the linearized plasmid BC1:KS containing the rat BC1 gene was performed using Bal 31 nuclease according to the supplier's recommendations (BRL, Gaithersburg, MD). The nuclease treated DNA was then size separated on an agarose gel and a gel slice

was removed which corresponded to the deletions of interest. The purified DNA (Geneclean, Bio 101) was then end-repaired using Klenow and T4 DNA polymerases, as previously described (Martignetti et al., 1992) and then subcloned into pBluescript:KS+ (Stratagene) predigested with EcoR V and BamH I restriction enzymes. All DNA used as transcription templates were CsCl purified via ultracentrifugation and then dialysed extensively against TE (10 mM: 0.1 mM, pH 8.0).

For the 3' deletion mutants the plasmid BC1:KS was digested with either Apa I or EcoR V, which releases both BC1 3' flanking region and part of the polylinker sequence, and then self-ligated.

Mutagenesis.

Site directed mutagenesis was performed as per Kunkel (1985) where single-stranded templates were grown in the E. coli CJ236 strain (dut$^-$ ung$^-$ double mutant). The oligonucleotides used were:

BC1: -(A): 5'-CTAGGCAAGCGCTCTAATCCCCAACCCCGA-3'

BC1: Δ A: 5'-TTTTTCGGAGCTGACTAAGCTTGCCAGGGCCTTGC-3'

BC1: -(B): 5'-TTTTTTCGGAGCTGACAGGGCCTTGCGCTT-3'

BC1: Δ B: 5'-GGCAAGCGCTCTACCTAAGCTTGAAATCCCCAACC-3'

BC1: B box: 5'-GGAGCTGAGGACTCGAACCCAG-3'

BC1: +5 /tRNA: 5'-ATCCTGACAGCAGGTCTCGTTAGCTC-3'

BC1:Δ TATA: 5'-AAATACTTCAGTTTTGCTCTCGAGACTGGCAGC-3'

BC1 Δ PSE L: 5'-GTACCATCTGATACTCGGATCCGTATGAAATAC-3'

BC1 Δ PSE R: 5'-TACTTGACTGTGGATCCGCTACTTCAGTTTTGC-3'.

The BC1: -(A + B) construct was prepared sequentially. All mutations were verified by DNA sequence analysis as described under DNA Sequencing.

DNA Sequencing

All clones were manually sequenced using the dideoxy sequencing method as previously described (Bankier et al, 199x) on single-stranded templates. These were prepared from Bluescript vectors and M13 K07 helper phage.

Transcription extracts.

Whole cell brain extract was prepared from young (150g), male Sprague-Dawley rats. Animals were decapitated and cortex was placed on a cold stage, diced and immediately placed into 3 volumes of ice cold buffer I (20 mM HEPES (pH 7.9 at 25° C), 10 mM $NaMoO_4$, 2 mM EDTA, 2 mM EGTA, 0.3 mM Spermine, 1 mM Spermidine, and 2 mM DTT, .5 mM PMSF, 5 μg/ml Aprotinin which were always prepared fresh and added before use). The tissue was kept in ice until all of the sample was collected. The usual total wet volume of cortex tissue collected was approximately 10 mls. At this point all reactions were performed in a cold room. After 15 minutes, the cells were lysed by hand in a Dounce homogenizer (always kept in ice) using a "B" pestle, until the mixture looked homogeneous in consistency. Four volumes of buffer II (20 mM HEPES (pH 7.9 at 25° C), 5 mM $NaMoO_4$, 1 mM $MgCl_2$, 25% Sucrose, 50% Glycerol, 2 mM DTT, .5 mM PMSF, and 5 μg/ml Aprotinin) was added and gently mixed. One volume of a saturated and neutralised solution of ammonium sulfate was added slowly with continued mixing. After 45 minutes on ice, the extract was poured into swinging bucket tubes and spun at 175,000 g_{av} for 3.5 hrs at 2° C. The topmost solid layer, containing lipids, was discarded and the supernatant collected without

disturbing the pellet. Ammonium sulfate, 0.35 g/ml of supernatant, was slowly added to the gently mixing suspension. One ul of 1M NaOH /g of ammonium sulfate was then added, and the suspension was allowed to mix for an additional 30 min. The precipitated proteins were collected by centrifugation at 20,000xg for 20 min. at 2° C. The supernatant was drained and the resultant protein pellet resuspended in a minimum volume of buffer III (20 mM HEPES (pH 7.9 at 25° C), 100 mM KCl, 10 mM $NaMO_4$, .1 mM EDTA, .1 mM EGTA, 18% Glycerol, and 1 mM DTT and .1 mM PMSF). The suspension was dialyzed twice against >250 fold excess of buffer III for 4-8 hours. Aliquots were snap frozen in liquid N_2 and stored at -80° C. The extracts have remained transcriptionally active for greater than 1 year.

HeLa cell extracts were prepared from HeLa S3 tissue culture cells by the method of Dignam et al (1983) and were the generous gifts of Shona Murphy and Jim Bieker.

In Vitro Transcription reactions.

Typically, the reaction mixes contained: 50% whole cell extract, 0.4 mM ATP, CTP, UTP, 20μM GTP, 10 μCi (alpha ^{32}P) GTP (800 Ci/mM), 10 mM Tris pH 7.9, 1 mM DTT, 5 mM creatine, 1 μg/ml α-amanitin, .5 U RNAsin (Promega) and 10 - 25 μg/ml template DNA in a total volume of 25 μl. For reactions employing HeLa cell extracts, 30% extract was used. All reactions were incubated at 30° C for 1 hr. The reactions were then digested with DNAse I (100μg/ml) for 5 min. and 125 μl solution of 0.2M NaAc, 0.5% SDS, 40 μg/ml yeast tRNA were added. These were extracted once with half volume phenol, quarter volume chloroform:isoamyl alcohol (24:1) and precipitated with two volumes of EtOH at -20° C for at

least 1 hr. The precipitated RNA was then resuspended in 10µl 99% formamide, 5 mM EDTA, 0.1% bromophenol blue, 0.2% xylene cyanol, heat denatured at 70° C for 5 minutes and electrophoresed on 6% acrylamide, 7M urea gels.

All reactions were repeated at least three times and several of the templates were prepared in duplicate to ensure that neither inhibition nor enhancement seen in the transcription reactions was due to template preparation. For the 5' deletion constructs, the estimates of transcriptional efficiency were obtained by densitometric scanning of several autoradiographs. These determinations were kindly performed by J. Berman.

REFERENCES

Aufiero, B., and Schneider, R. J. (1990). The hepatitis B virus X-gene product trans-activates both RNA polymerase II and III promoters. The EMBO J., 9, 497-504.

Bankier, A. T., Weston, K. M., and Barrell, B. G. (1987). Methods Enzymol., 155, 51-93.

Bazett-Jones, D. P., Yeckel, M., and Gottesfeld, J. M. (1985). Nuclear extracts from globin-synthesizing cells enhance globin transcription *in vitro*. Nature, 317, 824-828.

Bieker, J. J., Martin, P. L., and Roeder, R. G. (1985). Formation of a rate-limiting intermediate in 5S gene transcription. Cell, 40, 119-127.

Brown, D. D. (1982). How a simple animal gene works. The Harvey Lectures, 76, 27-44.

Carbon, P., Murgo, S., Ebel, J.-P., Krol, A., Tebb, G., and Mattaj, I. W. (1987). A common octamer motif binding protein is involved in the transcription of U6 snRNA by RNA polymerase III and U2 snRNA by RNA polymerase II. Cell, 1987, 71-79.

Ciliberto, G., Castagnoli, L., and Cortese, R. (1983). Transcription by RNA polymerase III. Current Topics in Developmental Biology, 18, 59-88.

Clarkson, S. G., Kurer, V. and Smith, H. O. (1978). Sequence organization of a cloned $tDNA^{Met}_1$ fragment from Xenopus laevis. Cell, 14, 713-724.

Comai, L., Tanese, N., and Tjian, R. (1992). The TATA-binding protein and associated factors are integral components of the RNA polymerase I transcription factor, SL1. Cell, 68, 965-976.

Cromlish, J. A., and Roeder, R. G. (1989). Human transcription factor IIIC (TFIIIC). J. Biol. Chem., 264, 18100-18109.

Daniels, G. R., and Deininger, P. L. (1985). Repeat sequence families derived from mammalian tRNA genes. Nature, 317, 819-822.

Das, G., Henning, D., Wright, D., and Reddy, R. (1988). Upstream regulatory elements are necessary for transcription of a U6 RNA gene by RNA polymerase III. EMBO J., 7, 503-512.

De Franco, D., Schmidt, O., and Soll, D. (1980). Two control regions for eucaryotic tRNA gene transcription. Proc. Natl. Acad. Sci. USA, 77, 3365-3368.

DeChiara, T. M., and Brosius, J. (1987). Neural BC1 RNA: cDNA clones reveal nonrepetitive sequence content. Proc. Natl. Acad. Sci. USA, 84, 2624-2628.

Dignam, J. D., Lebovitz, R. M., and Roeder, R. G. (1983). Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucl. Acids Res., 11, 1475-1489.

Dingermann, T., Burke, D. J., Sharp, S., Schaack, J., and Soll, D. (1982). The 5' flanking sequences of *Drosophila* $tRNA^{Arg}$ genes control their in vitro transcription in a *Drosophila* cell extract. The J. Biol. Chem., 257, 14738-14744.

Dingermann, T., Frank-Stoll, U., Werner, H., Wissmann, A., Hillen, W., Jacquet, M., and Marschalek, R. (1992). RNA polymerase III catalyzed transcription can be regulated in Saccharomyces cerevisiae by the bacterial tetracycline repressor-operator system. The EMBO J., 11, 1487-1492.

Folk, W. R. (1988). Changing direction in pol III transcription. Genes Devel., 2, 373-375.

Fowlkes, D. M., and Shenk, T. (1980). Transcriptional control regions of the Adenovirus VAI RNA gene. Cell, 22, 405-413.

Fradkin, L. G., Yoshinaga, S. K., Berk, A. J., and Dasgupta, A. (1987). Inhibition of host cell RNA polymerase III-mediated transcription by poliovirus: inactivation of specific transcription factors. Mol. Cell. Bio., 7, 388-3887.

Gabrielsen, O. S., Marzouki, N., Ruet, A., Sentenac, A., and Fromageot, P. (1989). Two polypeptide chains in yeast transcription factor τ interact with DNA. J. Biol. Chem., 264, 7505-7511.

Gabrielsen, O. S., and Sentenac, A. (1991). RNA polymerase III (C) and its transcription factors. Trends in Biochem., 16, 412-416.

Galli, G., Hofstetter, H., and Birnstiel, M. L. (1981). Two conserved sequence blocks within eukaryotic tRNA genes are major promoter elements. Nature, 294, 626-631.

Garcia, A. D., O'Connell, A. M., and Sharp, S. J. (1987). Formation of an active transcription complex in the *Drosophila melanogaster* 5S RNA gene is dependent on an upstream region. Mol. Cell. Bio., 7, 2046-2051.

Geiduschek, E. P., and Tocchini-Valentini, G. P. (1988). Transcription by RNA polymerase III. Ann. Rev. Biochem., 57, 873-914.

Gorski, K., Carneiro, M., and Schibler, U. (1986). Tissue-specific in vitro transcription from the mouse albumin promoter. Cell, 47, 767-776.

Hipskind, R. A., and Clarkson, S. G. (1983). 5' flanking sequences that inhibit *in vitro* transcription of a *Xenopus laevis* tRNA gene. Cell, 34, 881-890.

Hoeffler, W. K., Kovelman, R., and Roeder, R. G. (1988). Cell, 53, 907-920.

Hoeffler, W. K., and Roeder, R. G. (1985). Enhancement of RNA polymerase III transcription by the E1A gene product of adenovirus. Cell, 41, 955-963.

Howe, J. G., and Shu, M.-D. (1989). Epstein-Barr Virus small RNA (EBER) genes: unique transcription units that combine RNA polymerase II and III promoter elements. Cell, 57, 825-834.

Johnson, D. L., and Wilson, S. L. (1989). Identification of a 150 kilodalton polypeptide that copurifies with yeast TFIIIC and binds specifically to tRNA genes. Mol. Cell. Biol., 9, 2018-2024.

Johnson, J. D., and Raymond, G. J. (1984). Three regions of a yeast $tRNA^{Leu}_3$ gene promote RNA polymerase III transcription. The J. Biol. Chem., 259, 5990-5994.

Kassavetis, G. A., Braun, B. R., Nguyen, L. H., and Geiduschek, E. P. (1990). S. cerevisiae TFIIIB is the transcription initiation factor proper of RNA polymerase III, while TFIIIA and TFIIIC are assembly factors. Cell, 60, 235-240.

Kassavetis, G. A., Riggs, D. L., Negri, R., Nguyen, L. H., and Geiduschek, E. P. (1989). Transcription factor IIIB generates extended DNA interactions in RNA polymerase III transcription complexes on tRNA genes. Mol. Cell. Biol., 9, 2551-2566.

Kleinert, H., and Benecke, B.-J. (1988). Transcription of human 7SK DNA *in vitro* and *in vivo* is exclusively controlled by an upstream promoter. Nucl. Acids Res., 16, 1319-1331.

Kressman, A., Hofstetter, H., DiCapua, E., Grosschedl, R., and Birnstiel, M. L. (1979). A tRNA gene of *Xenopus laevis* contains at least two sites promoting transcription. Nucl. Acids Res., 7, 1749-1763.

Kunkel, G. R. (1991). RNA polymerase III transcription of genes that lack internal control regions. Biochim. et Biophys. Acta, 1088, 1-9.

Kunkel, G. R., and Pederson, T. (1988). Upstream elements required for efficient transcription of a human U6 gene resemble those of U1 and U2 genes even though a different polymerase is used. Genes Dev., 2, 196-204.

Kunkel, T. A. (1985). Rapid and efficient site specific mutagenesis without phenotypic selection. Proc. Natl. Acad. Sci. USA, 82, 488-492.

Lassar, A. B., Martin, P. L., and Roeder, R. G. (1983). Transcription of class III genes: formation of preinitiation complexes. Science, 222, 740-748.

Lawrence, C. B., McDonnell, D. P., and Ramsey, W. J. (1985). Nucl. Acids Res., 13, 4239-4252.

Lobo, S. M., and Hernandez, N. (1989). A 7 bp mutation converts a human RNA polymerase II snRNA promoter into an RNA polymerase III promoter. Cell, 58, 55-67.

Lobo, S. M., Lister, J., Sullivan, M. L., and Hernandez, N. (1991). The cloned RNA polymerase II transcription factor IID selects RNA polymerase III to transcribe the human U6 gene in vitro. Genes Devel., 5, 1477-1489.

Makowski, D. R., Haas, R. A., Dolan, K. P., Grunberger, D. (1983). Molecular cloning, sequence analysis and in vitro expression of a rat tRNA gene cluster. Nucl. Acids. Res. 11, 8609-8624.

Manley, J. L. (1984). Transcription of eukaryotic genes in a whole cell extract. In B. D. Hames and S. J. Higgins (Eds.), Transcription and Translation: a practical approach (pp. 71-88). Washington DC: IRL Press.

Manley, J. L., Fire, A., Cano, A., Sharp, P. A., and Gefter, M. L. (1980). DNA-dependent transcription of adenovirus genes in a soluble whole-cell extract. Proc. Natl. Acad. Sci. USA, 77, 3855-3859.

Margottin, F., Dujardin, G., Gerard, M., Egly, J.-M., Huet, J., and Sentenac, A. (1991). Participation of the TATA factor in transcription of the yeast U6 gene by RNA polymerase C. Science, 251, 424-426.

Martignetti, J. A., and Brosius, J. (1992a). A neural specific RNA polymerase III product encoded by a monomeric Alu element. Submitted.

Martignetti, J. A., and et al. (1992b). The rodent BC1 RNA genes. Manuscript in preparation.

Mattaj, I. W., Dathan, N. A., Parry, H. D., Carbon, P., and Krol, A. (1988). Changing the RNA polymerase specificity of U snRNA gene promoters. Cell, 55, 435-442.

McKinnon, R. D., Shinnick, T. M., and Sutcliffe, J. G. (1986). The neuronal identifier element is a *cis*-acting positive regulator of gene expression. Proc. Natl. Acad. Sci. USA, 83, 3751-3755.

Meza, L., Araya, A., Leon, G., Krauskopf, M., Siddiqui, M. A. Q., and Garel, J. P. (1977). Specific alanine tRNA species associated with fibroin synthesis in the posterior silkgland of Bombyx mori. FEBS Letters, 77, 255-260.

Morton, D. G., and Sprague, K. U. (1984). In vitro transcription of a silkworm 5S RNA gene requires an upstream signal. Proc. Natl. Acad. Sci. USA, 81, 5519-5522.

Murphy, M. H., and Baralle, F. E. (1983). Directed semisynthetic point mutational analysis of an RNA polymerase III promoter. Nucl. Acids Res., 11, 7695-7700.

Murphy, S., Di Liegro, C., and Melli, M. (1987). The in vitro transcription of the 7SK RNA gene by RNA polymerase III is dependent only on the presence of an upstream promoter. Cell, 51, 81-87.

Murphy, S., Moorefield, B., and Pieler, T. (1989). Common mechanisms of promoter recognition by RNA polymerase II and III. Trends in Genet., 5, 122-126.

Palmer, J. M., and Folk, W. R. (1990). Unraveling the complexities of transcription by RNA polymerase III. Trends in Biochem., 15, 300-304.

Parry, H. D., and Mattaj, I. W. (1990). Positive and negative functional interactions between promoter elements from different classes of RNA polymerase III-transcribed genes. EMBO J., 9, 1097-1104.

Parsons, M. C., and Weil, P. A. (1990). Purification and characterization of Saccharomyces cerevisiae transcription factor TFIIIC. J. Biol. Chem., 265, 5095-5103.

Reynolds, W. F., and Johnson, D. L. (1992). Differential expression of oocyte-type class III genes with fraction TFIIIC from immature or mature oocytes. Mol. Cell. Biol., 12, 946-953.

Sakamoto, K., and Okada, N. (1985). 5-methylcytidylic modification of in vitro transcript from the rat identifier sequence; evidence that the transcript forms a tRNA-like structure. Nucl. Acids Res., 13, 7195-7206.

Schaack, J., Sharp, S., Dingermann, T., and Soll, D. (1983). Transcription of eukaryotic tRNA genes *in vitro*. II. Formation of stable complexes. J. Biol. Chem., 258, 2447-2453.

Scholer, H. R., and Gruss, P. (1985). Cell type-specific transcriptional enhancement *in vitro* requires the presence of *trans*-acting factors. EMBO J., 4, 3005-3013.

Schwartz, L. B., Sklar, V. E. F., Jaehning, J. A., Weinman, R., and Roeder, R. G. (1974). Isolation and partial characterization of the multiple forms of DNA dependent RNA polymerase in the mouse myeloma MOPC-315. J. Biol. Chem., 249, 5889-5897.

Segall, J., Matsui, T., and Roeder, R. G. (1980). Multiple factors are required for the accurate transcription of purified genes by RNA polymerase III. J. Biol. Chem., 255, 11986-11991.

Selker, E. U., Morzycka-Wroblewska, E., Stevens, J. N., and Metzenberg, R. L. (1986). An upstream signal is required for *in vitro* transcription of *Neurospora* 5S RNA genes. Mol. Gen. Genet., 205, 189-192.

Sharp, P. A. (1992). TATA-binding protein is a classless factor. Cell, 68, 819-821.

Sharp, S., DeFranco, D., Dingermann, T., Farrell, P., and Soll, D. (1981). Internal control regions for transcription of eukaryotic tRNA genes. Proc. Natl. Acad. Sci. USA, 78, 6657-6661.

Sharp, S. J., Schaack, J., Cooley, L., Burke, D. J., and Soll, D. (1984). Structure and transcription of eukaryotic transfer RNA genes. CRC Crit. Rev. Biochem., 19, 107-144.

Simmen, K. A., Bernues, J., Parry, H. D., Stunnenberg, H. G., Berkenstam, A., Cavallini, B., Egly, J.-M., and Mattaj, I. W. (1991). TFIID is required for in vitro transcription of the human U6 gene by RNA polymerase III. EMBO J., 10, 1853-1862.

Sollner-Webb, B. (1988). Surprises in polymerase III transcription. Cell, 52, 153-154.

Sprague, K. U., Hagenbuchle, O., and Zuniga, M. C. (1977). The nucleotide sequence of two silk gland alanine tRNAs: implications for fibroin synthesis and for initiator tRNA structure. Cell, 11, 561-570.

Sprague, K. U., Larson, D., and Morton, D. (1980). 5' flanking sequence signals are required for activity of silkworm alanine tRNA genes in homologous *in vitro* transcription systems. Cell, 22, 171-178.

Sugden, B., and Keller, W. (1973). Mammalian DNA dependent RNA polymerases part 1: purification and properties of an alpha amanitin

sensitive RNA polymerase and stimulatory factors from HeLa and KB cells. J. Biol. Chem., 248, 3777-3788.

Sutcliffe, J. G., Milner, R. J., Bloom, F. E., and Lerner, R. A. (1982). Common 82-nucleotide sequence unique to brain RNA. Proc. Natl. Acad. Sci. USA, 79, 4942-4946.

Sutcliffe, J. G., Milner, R. J., Gottesfeld, J. M., and Reynolds, W. (1984). Control of neuronal gene expression. Science, 225, 1308-1315.

Tiedge, H., Fremeau Jr., R. T., Weinstock, P. H., Arancio, O., and Brosius, J. (1991). Dendritic location of neural BC1 RNA. Proc. Natl. Acad. Sci. USA, 88, 2093-2097.

Traboni, C., Ciliberto, G., and Cortese, R. (1984). Mutations in box B of the promoter of a eucaryotic $tRNA^{Pro}$ gene affect aate of transcription, processing, and stability of the transcripts. Cell, 36, 179-187.

Tsuda, M., and Suzuki, Y. (1981). Faithful transcription initiation of fibroin gene in a homologous cell-free system reveals an enhancing effect of 5' flanking sequence far upstream. Cell, 27, 175-182.

Ullu, E., and Weiner, A. M. (1985). Upstream sequences modulate the internal promoter of the human 7SL RNA gene. Nature, 318, 371-375.

Waibel, F., and Filipowicz, W. (1990). RNA-polymerase specificity of transcription of Arabidopsis U snRNA genes determined by promoter element spacing. Nature, 346, 199-202.

Watson, J. B., and Sutcliffe, J. G. (1987). Primate brain-specific cytoplasmic transcript of the Alu repeat family. Mol. Cell. Biol., 7(9), 3324-3327.

White, R. J., Stott, D., and Rigby, P. W. J. (1990). Regulation of RNA polymerase III transcription in response to Simian virus 40 transformation. The EMBO J., 9, 3713-3721.

Wu, G.-J. (1978). Adenovirus DNA-directed transcription of 5.5S RNA *in vitro*. Proc. Natl. Acad. Sci. USA, 75, 2175-2179.

Yoshinaga, S., Dean, N., Han, M., and Berk, A. J. (1986). EMBO J., 5, 343-354.

Figure 1. Effect of $MgCl_2$ and temperature on the *in vitro* transcription reaction. Extract preparation, template assay and product analysis were as described in Experimental Procedures except for the variations described in (A) and (B). For each condition tested, a $tRNA^{Met}$ ( ) and BC1 RNA (BC1:KS) gene containing template were assayed and the final concentration of each template was 25 μg/ml. Template BC1:KS in this and subsequent experiments refers to the -429 to +1453 template shown in Figure 2B. The relative position of the $tRNA^{Met}$ and BC1 RNA products are indicated. (A). Effect of increasing the final $MgCl_2$ concentration in the reaction mixture. Lane 1, 1 mM $MgCl_2$; Lane 2, 3 mM $MgCl_2$; Lane 3, 5 mM $MgCl_2$; Lane 4, 9 mM $MgCl_2$. (B) Effect of temperature during incubation. Transcription reactions were prepared on ice and then incubated for 60 minutes at the different temperatures indicated. Lane 1, 20° C; Lane 2, 25° C; Lane 3, 30° C; Lane 4, 37° C; Lane 5, 42° C. Different preparations of whole cell extracts were used in A and B.

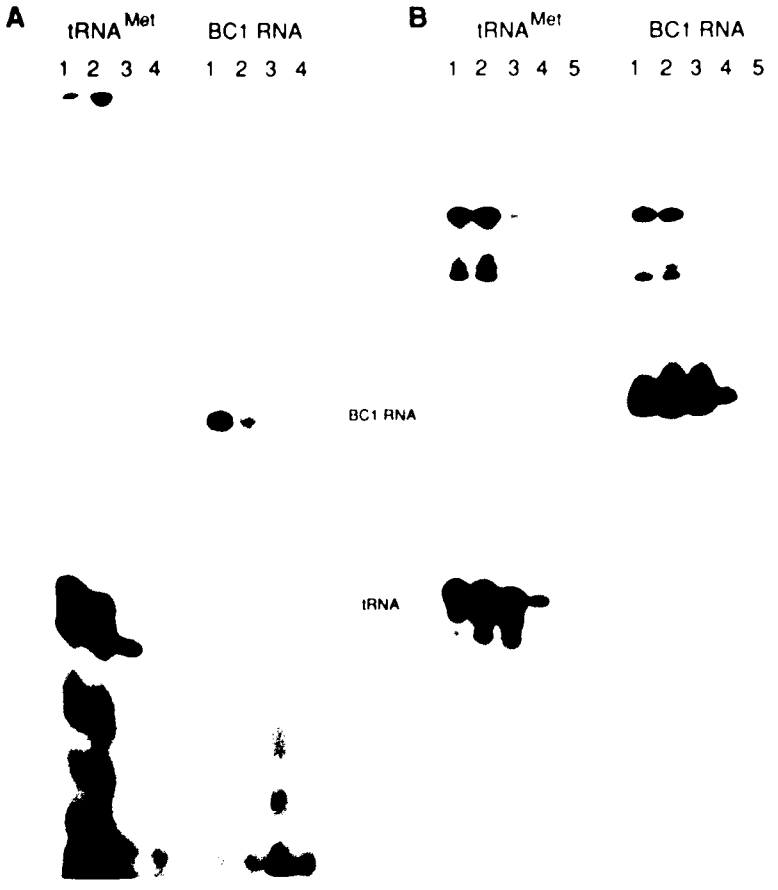

A
tRNA^Met
BC1 RNA
1 2 3 4
1 2 3 4
B
tRNA^Met
BC1 RNA
1 2 3 4 5
1 2 3 4 5
BC1 RNA
tRNA

Figure 2. The sequence of the BC1 RNA upstream and coding region. (A) The sequence of the rat BC1 gene (Martignetti et al., 1992) is shown from position -423, contiguous to a Sac I site, not shown, to position +154, at a stretch of 5 T's, believed to be represent the termination signal. The *in vivo* start and termination sites () have been highlighted by a thick arrow and a capped vertical line, respectively. (B) Diagram of the rat BC1 RNA gene (BC1: KS) representing the relative positions of the promoter elements discussed in the text.

A

```
-423 TCATTTTCAAAGAGCACCATGTTAGAGTTAAGGATTCATTTGTATATATGAAATTCAGATTGGAGGTAGAATCCCTGGAG
                                          OCTA
-343 TGTGTGTGTATGTGTGTGTGTGTGTTCACTTAGCCCTCCAATCAATTATATTGTGTCTCTGGTCTCTTTCATACAGAAGC

-263 TGATCAACTTGTTAGCAATTGTTGTCTAAGGAATTTACTGTACTTTGGAATACTTTTTTAAACTATATTTTGGAAGGTAT

-183 CTCTGATGTAAATGAGAAAGCTAACTTCAGTTTCTGCTTTTTGGAATGTGGGTGCCTATGGGATTGGAGCGCTTGGACAA
       OCTA
-103 AGTGGCTCCTCCTGCCGGCCAGCCCTTCGGTACCATCTGATACTTGACTGTGTATGAAATACTTCAGTTTTGCTTTAAAT
                                                PSE                              TATA
 -23 ACTGGCAGCAAGAGCTAACGTTCGGGGTTGGGGATTTAGCTCAGTGGTAGAGCGCTTGCCTAGCAAGCGCAAGGCCCTGG
                                          A box                                    ●
     GTTCGGTCCTCAGCTCCG(A)22GACAAAATAACAAAAAGACCAAAAAAAAAACAAGGTAACTGGCACACACAACCTTTTT
       B box
```

B

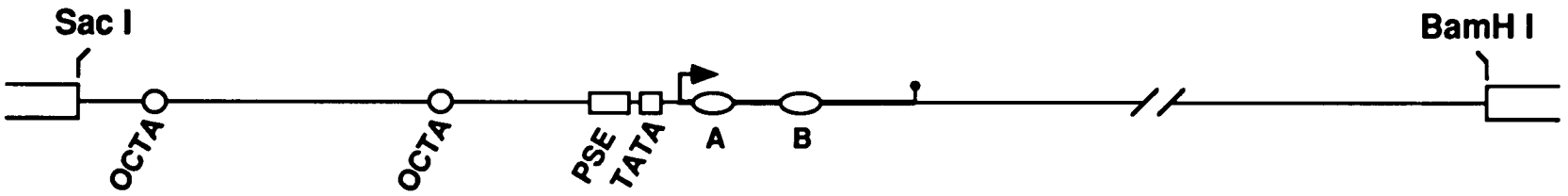


Figure 3. The BC1 intragenic promoter sequences are required for *in vitro* transcription. (A) Deletion of the intragenic BC1 A and B boxes as assayed in whole cell cortex extracts. The series of intragenic deletions used is shown in I where the A and B boxes are represented by open ovals. (B) Effect of mutation of the intragenic BC1 A and B boxes as assayed in cortex extracts. The series of mutations used is shown in II where the mutated sequence is represented by a shaded rectangle. (C) and (D) are as (A) and (B), respectively, except that the templates were assayed in whole cell HeLa extracts. The BC1 RNA transcripts are indicated by filled arrows. The assays performed in the whole cell cortex extracts were exposed to film for 12 hours, those from HeLa extract, 2 hours.

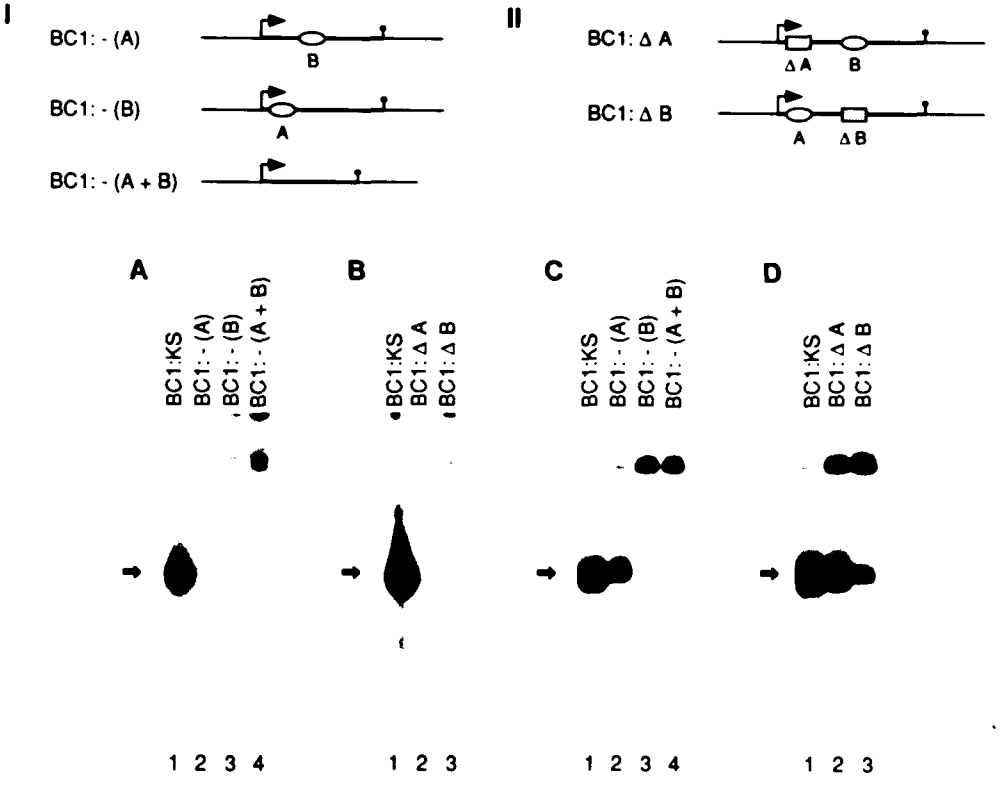
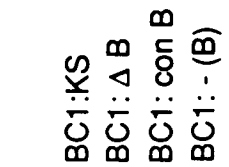
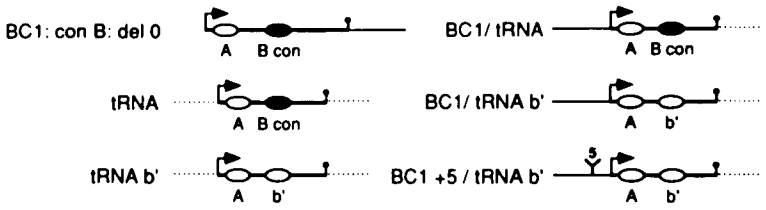

I
BC1: - (A)
B
BC1: - (B)
A
BC1: - (A + B)
II
BC1: Δ A
ΔA
B
BC1: Δ B
A
ΔB
A
BC1:KS
BC1: - (A)
BC1: - (B)
BC1: - (A + B)
1 2 3 4
B
BC1:KS
BC1: Δ A
BC1: Δ B
1 2 3
C
BC1:KS
BC1: - (A)
BC1: - (B)
BC1: - (A + B)
1 2 3 4
D
BC1:KS
BC1: Δ A
BC1: Δ B
1 2 3

Figure 4. The BC1 b' box and a consensus B box are not mutually interconvertable. The mutation of the BC1 B box, b' box, to equal the consensus B box sequence (BC1: con B) results in decreased levels of transcription as compared to wild type (BC1:KS) but greater than if the BC1 b' box is mutated to a non-related sequence (BC1: ΔB) or completely deleted (BC1: -(B)). The BC1 RNA transcripts are indicated by a filled arrow. The assay was performed using the whole cell cortex extract.



1 2 3 4

Figure 5. The effect of interaction between intragenic and extragenic sequences on *in vitro* synthesis. (A) Comparison of BC1 RNA gene templates containing the wild type BC1 b' box, a consensus B box sequence replacing the b' box, as in Figure 4, and a B box construction with all BC1 5' flanking sequences removed (BC1: con B: del 0). (B) Comparison of different rat $tRNA^{Leu}$ () gene constructions. (C) Comparison of the effect of promoter spacing in hybrid BC1/$tRNA^{Leu}$ constructions. The exposure time of the transcription products in (B) was approximately twice that of the products in (C). The BC1 RNA transcripts are indicated by filled arrows, the $tRNA^{Leu}$ transcripts by open arrows. All assays were performed using the whole cell cortex extracts.


BC1: con B: del 0
A B con
BC1/ tRNA
A B con
tRNA
A B con
BC1/ tRNA b'
A b'
tRNA b'
A b'
BC1 +5 / tRNA b'
5
A b'

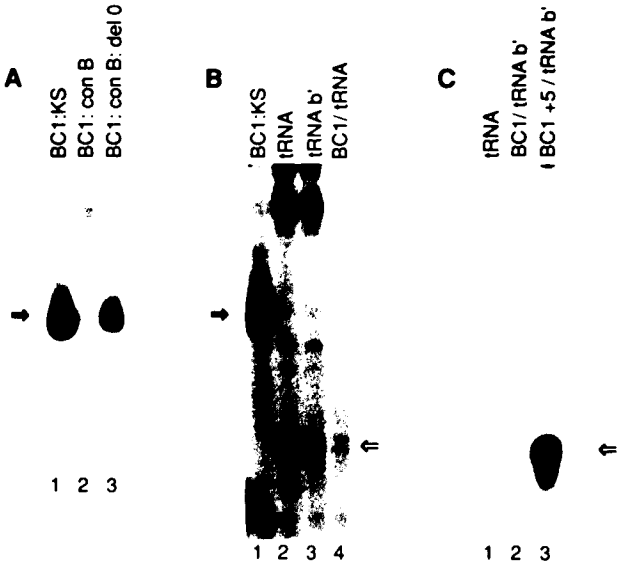

A
BC1:KS
BC1: con B
BC1: con B: del 0
1 2 3
B
BC1:KS
tRNA
tRNA b'
BC1/ tRNA
1 2 3 4
C
tRNA
BC1/ tRNA b'
BC1 +5 / tRNA b'
1 2 3

Figure 6. Mapping the BC1 RNA gene upstream sequences that regulate *in vitro* transcription. (A) Diagram of the deletions used where symbols represent the promoter elements highlighted in Figure 2 and numbers on the left side represent nucleotide position with respect to the BC1 RNA start site. (B) Upstream sequences are essential for BC1 RNA synthesis when assayed in the whole cell cortex extracts. (C) Transcription of the BC1 RNA gene in HeLa cell derived extracts does not require upstream sequences. Exposure times were as for Figure 3. The BC1 RNA transcripts are indicated by filled arrows.

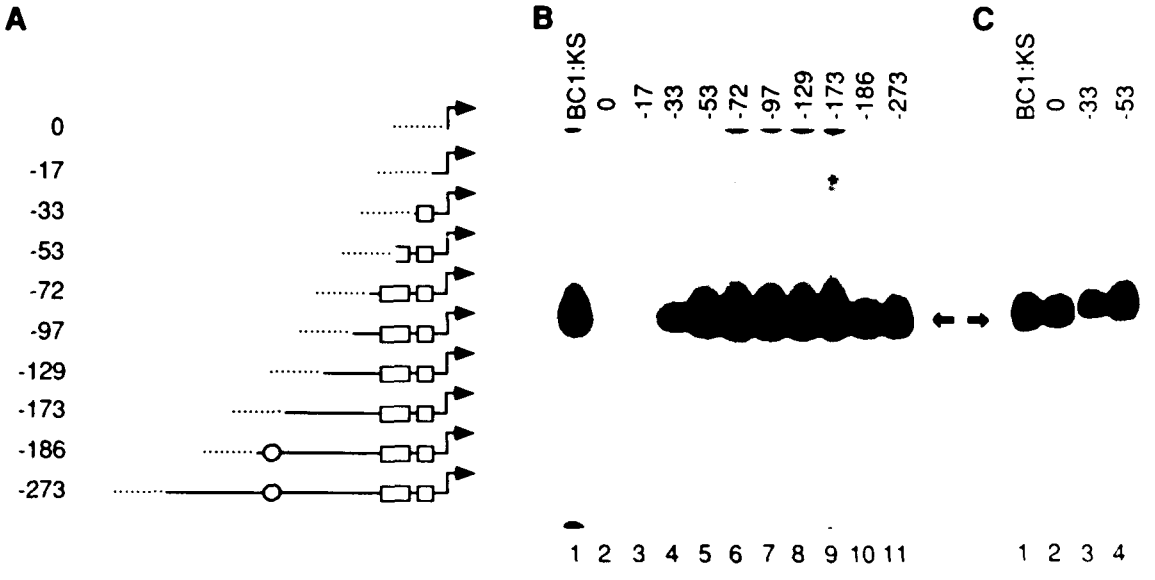
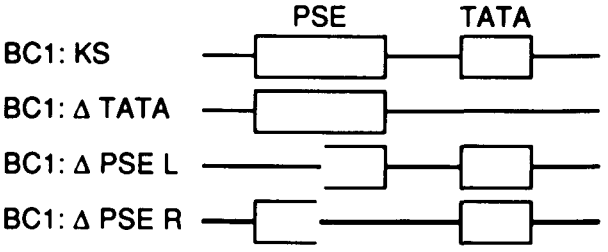

A
0
-17
-33
-53
-72
-97
-129
-173
-186
-273
B
BC1:KS
0
-17
-33
-53
-72
-97
-129
-173
-186
-273
1 2 3 4 5 6 7 8 9 10 11
C
BC1:KS
0
-33
-53
1 2 3 4

Figure 7. The BC1 TATA and PSE elements are functional in regulating BC1 RNA synthesis. (A) Dependance on the TATA and PSE left half for maximum *in vitro* transcription when assayed in the whole cell cortex extracts. (B) Mutation of the TATA and PSE elements do not affect transcription of the BC1 RNA gene when assayed in HeLa derived extracts


PSE
TATA
BC1: KS
BC1: Δ TATA
BC1: Δ PSE L
BC1: Δ PSE R

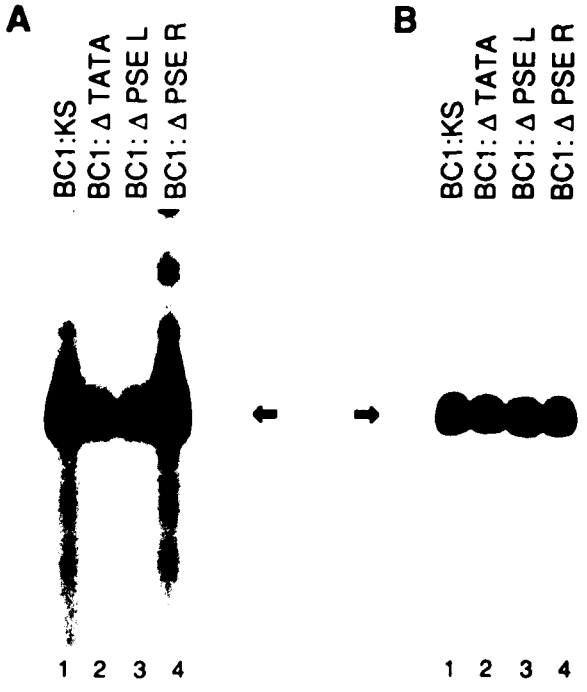

A
BC1:KS
BC1: Δ TATA
BC1: Δ PSE L
BC1: Δ PSE R
1 2 3 4
B
BC1:KS
BC1: Δ TATA
BC1: Δ PSE L
BC1: Δ PSE R
1 2 3 4

Figure 8. Schematic model for the interaction between intragenic and extragenic factors in pol III transcription.

(A) For most class III genes, formation of the initiation complex is dependent upon the binding of TFIIIC, represented by the complex of oval shaped factors, which then directs the binding of TFIIIB.

(B) The formation of an initiation complex, as shown in (A), and exemplified by a U6 gene/tRNA chimaeric construct (Parry and Mattaj, 1990) and a yeast $tRNA^{Glu}$/bacterial *tet* operator containing construct (Dingermann et al., 1992), is impeded by the presence of upstream factors which sterically inhibit TFIIIC from binding stably to its intragenic promoter.

(C) The BC1 RNA gene combines, as EBER2 RNA gene (Howe and Shu, 1989), intragenic and extragenic promoter sequences, which are effective in directing transcription. One method by which this could be achieved is if the TFIIIC binding to the intragenic region was not sterically inhibited by upstream binding factors. The BC1 b' box would direct the binding by a modified TFIIIC which is able to interact with upstream factors.

(D) The modified TFIIIC without the presence of upstream stabilizing factors is only weakly associated with the b' promoter and is inefficient at forming a pre-initiation complex.

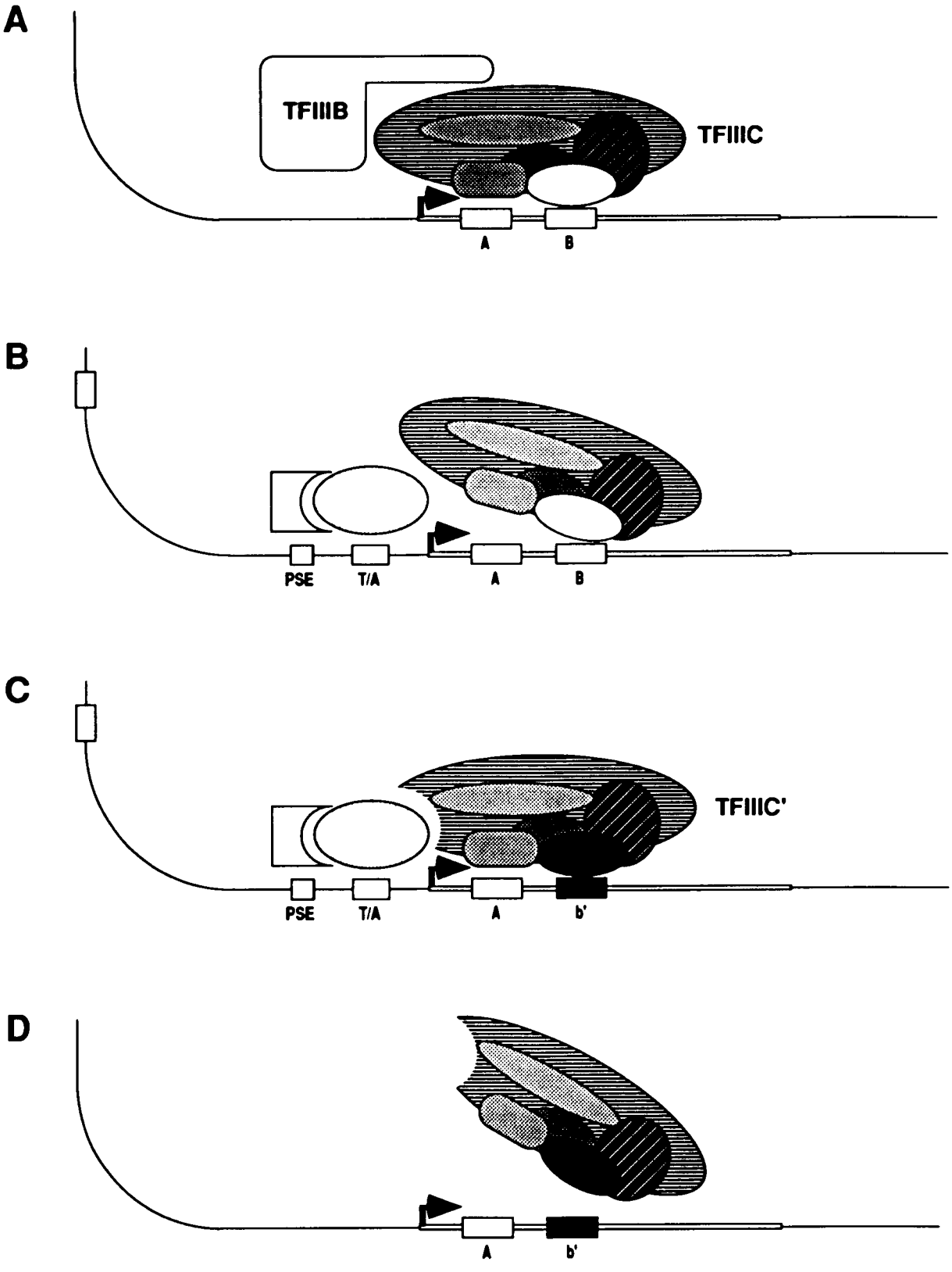

A
TFIIIB
TFIIIC
A
B
B
PSE
T/A
A
B
C
TFIIIC'
PSE
T/A
A
b'
D
A
b'

TABLE I

RNA	B box sequence
CONSENSUS	GGTTCGANNCC
BC1	GGTTCGGTCCT
BC1: con B	GGTTCGAGTCCT
tRNA$^{Leu\ 1}$	GGTTCGAATCC
tRNA$^{Leu\ b'}$	GGTTCGGTCCC
EBER 2 [2]	GGTGCTACCGA

Underlined nucleotides represent invariant positions

[1] as per Makowski et al., (1983)

[2] as per Howe and Shu (1989)

TABLE II

GTGACCGTGAGT(GT)RAAR$_{0-3}$TG C C TC		consensus PSE'
TTGACTGTGTAT	GAAA---TA	BC1:KS
agcttgaTGTAT	GAAA---TA	BC1:del -53
TCGGATCCGTAT	GAAA---TA	BC1:Δ PSE L
TTGACTGTGGAT	CCGC---TA	BC1:Δ PSE R
TGTCAAGAAGTG	TAAA---GT	7SK: PSE A[2]
TTGACCTAATGT	GCCC---TT	7SK: PSE B

'as per Kunkel (1991)

[2]as per Murphy et al., (1989)

Lower case letters denote vector supplied sequence.

CHAPTER 6

A Neural Specific RNA Polymerase III Product Encoded by a Monomeric Alu Element

John A. Martignetti

and

Jürgen Brosius

Fishberg Research Center for Neurobiology

Box 1065

Mount Sinai School of Medicine

1 Gustave L. Levy Pl.

New York, NY 10029

A driving force underlying evolution is the generation of new genes by the amplification of existing ones, as essential sequences are conserved while amplified copies, under no selective pressure, are free to mutate[1-3]. The 500,000 members of the Alu family of interspersed repetitive DNA constitutes approximately 5% of the genome[4] and are composed of dimeric structures retroposed from a small number of master genes[5-8]. These arose during primate evolution via the highly conserved and essential component of the signal recognition particle, 7SL RNA[9]. Here we report the identification of the first transcriptionally and transpositionally active Alu element. It encodes BC200 RNA, a neural specific, small cytoplasmic RNA expressed in the dendrites of a subpopulation of neuronal cells. Analysis of the BC200 gene defines its existence early in the genesis of the dimeric Alu family and the identification of BC200 RNA pseudogenes highlights its own role as a master gene. Furthermore, its product represents the first primate example of an RNA polymerase III (pol III) neural specific transcript. We suggest that BC200 RNA has acquired a useful role in the primate brain, confirming[10-13] that a repetitive DNA family member may confer a selective advantage to the organism.

BC200 RNA cDNA clones were isolated from a human brain small RNA library. From the generated sequence data an oligonucleotide probe was designed complementary to the 3' end which detects only a single band on Northern RNA blots corresponding to BC200 RNA (H. Tiedge, W. Chen and J. B., personal communication).

Using this probe, the BC200 RNA gene (α) and two BC200 pseudogenes (β and γ) were isolated from a human genomic DNA library (Fig. 1). Mapping studies localize each gene to an independent chromosome (A. Petronis, J. A. M., J. B. and J. Kennedy, manuscript in preparation).

Analysis of the BC200α gene sequence, as shown in figure 2, reveals that the BC200 RNA coding region consists of the left half of an Alu sequence, a central adenosine-rich tract and a 3' terminal non-repetitive 42 nucleotide (nt) sequence. Therefore, unlike the consensus primate Alu sequence which is dimeric in structure, composed of two similar and tandemly arranged monomeric repeats, the left (Lm) and right monomers (Rm)[14], primate BC200 RNA is in part encoded by an Alu Lm.

One theory of the origin of the dimeric Alu sequence maintains that it arose via a fusion of two independently derived monomeric elements[15]. A growing number of independent left and right Alu monomers or 'free left/right arms'[16-18] have been identified. Two lines of evidence suggest that the BC200α gene, the only identified active monomer, represents an early point of divergence or branchpoint in the evolution and amplification of Alu sequences.

First, the age of the BC200α gene was estimated by comparing its sequence to that of the different Alu subfamilies. The different subfamilies represent sequences amplified in waves from a discrete number of master genes within the past 60-70 million years[5-8]. Subfamilies are distinguishable from one another by sequence analysis: certain diagnostic sequence differences from an established consensus can be used to categorize and date a sequence of interest. The BC200α coding region matches all the diagnositic positions[19] of the left half of the

oldest subfamily, AluJ. It is therefore at least as old as the earliest dimeric sequence, calculated at 55 million years[20].

Second, structural analysis reveals that the BC200α gene shares sequence constraints with the 7SL RNA sequence not present in any known individual dimeric Alu sequence. As shown in figure 2, and allowing for the introduction of gaps, the BC200α gene and the AluJ Lm are equally homologous to 7SL RNA. Over the first 81 nts, BC200α and AluJ show 87% and 86% sequence homology to 7SL RNA, respectively. BC200 RNA and AluJ are more closely related, sharing 92% sequence homology. Most notably however, the BC200 RNA and 7SL RNA sequences share a high number of conserved CG doublets, despite the high transition rate of this dinucleotide pair[21]. While the consensus Alu sequence, an attempt at reconstructing the master element, also contains a similar number, it must be stressed that no known individual Alu J sequence possesses as many CG doublets[19].

Therefore, the BC200α gene represents an early member in the phylogeny of Alu sequences. Whether the BC200α gene contributed to the fusion event generating a dimeric Alu element is presently unclear. However, the isolation of two BC200 pseudogenes demonstrates that the BC200 RNA gene is itself a master gene for monomeric elements. Interestingly, the analysis of one of the pseudogenes shows that a BC200 generated monomer can generate a novel dimeric element.

The BC200β gene differs in only 3 positions (99% sequence homology) from the BC200 RNA coding region (Fig. 2). Surprisingly, an unrelated Alu Rm is found contiguous to 3' end of the BC200 RNA homology. The evolution of this chimaeric element was investigated by amplifying the corresponding orthologous locus from several primate

species (chimpanzee, gorilla, and orangutan). In comparison to human, these samples all yielded a similarly sized smaller product. The chimpanzee product, selected for sequencing and assumed representative of all three samples, lacks the BC200 RNA pseudogene but possesses a full-length, non-J type dimeric Alu element (Fig. 2). The compound BC200/Alu element was therefore generated by the replacement of the BC200α derived sequence for the younger Alu Lm. This occurred within the last 6 million years, after the human/chimpanzee divergence.

In marked contrast to the *in vivo* transcriptional silence of both the 500,000 Alu family members[22] and the several hundred 7SL RNA pseudogenes[23], BC200 RNA is the first transcriptionally active Alu element identified. Previous studies which identified the BC200 RNA in cynomolgus monkey[24] proves the RNA has been transcriptionally active in a neural specific manner for at least 20 million years.

As part of our investigation into studying its tissue specific transcriptional regulation, the RNA polymerase specificity of the BC200α gene was first determined. Using a HeLa whole cell extract, the synthesis of BC200 RNA is shown to be sensitive only to high concentrations of α-amanitin (figure 3). This suggests that BC200 RNA, expressed almost exclusively by a subpopulation of neurons in the primate nervous system, is transcribed by RNA pol III. This result is corroborated by the presence of a pol III termination signal consensus sequence, four or more T residues[25], found at the 3' end of the genomic sequence and the corresponding U residues seen in the terminal region of the mature RNA (H. Tiedge, W. Chen and J. B., personal communication).

A second set of transcription experiments deleted the BC200 RNA gene's 5' flanking sequence. This sequence contains the classic pol II promoter motif, a TATA box (figure 2, position -26). Deletion of the upstream region results in decreased transcriptional efficiency. This result may explain the finding that the BC200β and γ pseudogenes, with their internal promoter consensus sequences intact but lacking the BC200α gene's functional upstream sequences, are not transcribed in human cortex (H. Tiedge, W. Chen and J. B., personal communication).

These results parallel transcription studies of the 7SL RNA gene[26]. In addition to internal promoter elements, the first 37 nt upstream from the 7SL RNA transcription start site are required for efficient and accurate *in vitro* transcription. It has been hypothesized that the absence of 5' promoter-like flanking information near Alu sequences confines the majority of these retroposed sequences to transcriptional silence *in vivo* [15]. Interspersed repetitive elements have been characterized as "selfish DNA", capable of spreading throughout the genome, making no direct contribution to the fitness of the cell or organism, and whose only "function" seems to be self survival[27-29]. Unfortunately, the use of the anthropomorphic term belies the issue of biological consequences. That BC200 RNA has been recruited or "exapted"[11] to perform a function in the primate nervous system seems quite likely given its relative abundance and highly unusual dendritic localization, the evolutionary pressure maintained on the RNA coding sequence and that the BC200 RNA gene is transcriptionally active and represents the first known tissue-specific primate pol III transcribed gene.

Given the variety and nature of the functions performed by other small RNAs, such as 7SL RNA[30], the U snRNAs[31], RNAse P RNA[32] and TFIIIR[33], a distinct neural function could be defined by BC200 RNA. In light of growing evidence of the occurrence of translation in dendrites of nerve cells[34], and noting the BC200 RNA's dendritic localization and ability to bind two protein subunits of the signal recognition particle (D. Zopf, J. G. Cheng, J. B. and P. Walter, personal communication), it is tempting to speculate that BC200 RNA may be involved in regulating dendritic protein biosynthesis.

REFERENCES

1. Haldane, J. B. S. The Causes of Evolution (Longmans and Green, Oxford, 1932).

2. Muller, H. J. Genetics **17**, 237-252 (1935).

3. Ohno, S. Evolution by Gene Duplication (Springer, New York, 1970).

4. Schmid, C. W. & Jelinek, W. R. Science **216**, 1065-1070 (1982).

5. Deininger, P. L. & Slagel, V. K. Molec. cell. Biol **8**, 4566-4569 (1988).

6. Britten, R. J., Baron, W. F., Stout, D. B. & Davidson, E. H. Proc. natn. Acad. Sci. U.S.A.. **85**, 4770-4774 (1988).

7. Jurka, J. & Smith, T. Proc. natn. Acad. Sci. U.S.A. **85**, 4775-4778 (1988).

8. Deininger, P. L., Batzer, M. A., Hutchison III, C. A. & Edgell, M. H. Trends Genet., in press.

9. Ullu, E. & Tschudi, C. Nature **312**, 171-172 (1984).

10. Jacob, F. in Evolution from Molecules to Men. , (ed Bendall, D.S.) 131-144 (New York, Cambridge University Press, 1982).

11. Gould, S.J. & Vrba, E.S. Paleobio. **8**, 4-15 (1982).

12. Rogers, J. Nature **317**, 765-766 (1985).

13. Brosius, J. Science **251**, 753 (1991).

14. Deininger, P. L., Jolly, D. J. , Rubin, C. M., Friedmann, T. & Schmid, C. W. J. molec. Biol. **151**, 17-33 (1981).

15. Weiner, A. M., Deininger, P. L., & Efstratiadis, A. A. Rev. Biochem. **55**, 631-661 (1986).

16. Quentin, Y. J. molec. Evol. **27**, 194-202 (1988).

17. Jurka, J. & Zuckerkandl, E. J. molec. Evol. **33**, 49-56 (1991).

18. Quentin, Y. Nucleic Acids Res. **20**, 487-493(1992).

19. Jurka, J. & Milosavljevic, A. J. molec. Evol. **32**, 105-121 (1991).

20. Labuda, D. & Striker, G. Nucl. Acids Res. **17**, 2477-2491 (1989).

21. Bird, P. A. Trends Genet. **3**, 342-347 (1987).

22. Paulson, K. E., & Schmid, C.W. Nucl. Acids Res. **14**, 6145-6158 (1986).

23. Ullu, E. & Weiner, A. M. EMBO J., **3**, 3303-3310 (1984).

24. Watson, J. B. & Sutcliffe, J. G. Molec. cell. Biol. **7**, 3324-3327 (1987).

25. Bogenhagen, D. F. & Brown, D. D. Cell **24**, 261-270 (1981).

26. Ullu, E. & Weiner, A. M. Nature **318**, 371-374 (1985).

27. Dawkins, R. The Selfish Gene. (Oxford, Oxford Univ. Press, 1976).

28. Doolittle, W. F., & Sapienza, C. Nature **284**, 601-603 (1980).

29. Orgel, L. E., & Crick, F. H. C. Nature, **284**, 604-607 (1980).

30. Walter, P., & Blobel, G. Nature, **299**, 691-698 (1982).

31. Reddy, R. & Busch, H. In Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles (ed Birnstiel, M. L.) 1-37 (Berlin, Springer-Verlag, 1988).

32. Altman, S., Gold, H. A. & Bartiewicz, M. In Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles, (ed Birnstiel, M. L.) 183-195 (Berlin, Springer-Verlag, 1988).

33. Young, L. S., et al., Science, **252**, 542-546 (1991).

34. Banker, G. A, & Steward, O. Trends Neurosci., 15, 180-186 (1992).

35. Bankier, A. T., Weston, K. M. & Barrell, B. G. Methods Enzymol., **155**, 51-93 (1987).

36. Devereux, J., Haeberli, P. & Smithies, O. Nucl. Acids Res., **12**, 387-395 (1984).

37. Shen, M. R., Batzer, M.A. & Deininger, P. L. J. molec. Evol. **33**, 311-320 (1991).

38. Dignam, J. D., Lebovitz, R. M. & Roeder, R. G. Nucl. Acids Res. **11**, 1475-1489 (1983).

Figure 1. Structure of the human BC200α, β and γ genes (**A - C**, respectively). The BC200 RNA gene and the direction of transcription have been highlighted by a filled arrow, while the pseudogenes are indicated by shaded arrows. Alu elements and their orientation, with respect to their internal RNA pol III promoter, are indicated by the duplicated open arrow figures, representing the two monomer halves, while L1 elements are represented by open boxes with their orientation indicated by arrows. All clones were sequenced using the dideoxy chain termination method[35] and analysis was performed using the GCG suite of programs (version7.0)[36] made available by the Department of Biomathematics, Mount Sinai School of Medicine.

A
EcoR I
Alu
Alu
BC200α
L1
Alu
EcoR I
0
1000
2000
3000
4000
C
Taq I
Sal I
L1
BC200γ
0
1000
B
Taq I
Pst I
Alu
Alu
Pst I
Taq I
BC200β/Alu
Pst I
0
1000
2000
3000
4000
5000
6000

Figure 2. Sequence alignment of the BC200 RNA coding region. 7SL, human 7SL RNA[26]; Alu/LM, the left half of the Alu J consensus sequence[20]; α, BC200α gene; β, BC200β/Alu element; γ, BC200γ gene; β/CZ, β pseudogene locus amplified from chimpanzee; and, Alu/CS, the Catarrhine Specific Alu consensus sequence[37]. Sequences were initially aligned using the program PILEUP within the GCG suite of programs as in ref. 36 and then further aligned manually. The coding sequence of the BC200α gene has been underlined and the numbering above the sequences uses the presumed start site of the RNA as a reference. The flanking direct repeats of the different Alu consensus sequences are indicated by "*" while those of the BC200β gene and the orthologous chimpanzee locus are notated by lower case letters. The '>' symbol represents the beginning of the BC200γ gene which does not contain sequence homologous to the LINE 1 element (discussed in text). The square brackets, "[ ]", represent a sequence gap, such that position 267 of the 7SL sequence is the next position in the alignment. The parallel lines, "//" found at the end of sequences demarcate the ends of the left monomers. Sequence differences relative to the BC200α RNA coding sequence have been underlined, except for expansion of homopolymeric regions, and spaces have been introduced to maximize homology. The two nucleotide insertion present in the human BC200 sequences (positions 49/50) have been acquired in the past 20 million years, since they are not present in the rhesus and cynomolgus monkey BC200 RNA sequences. Numbering is based on comparison with the BC200 RNA coding sequence such that the sequence gap following position 37 is not counted as a space.

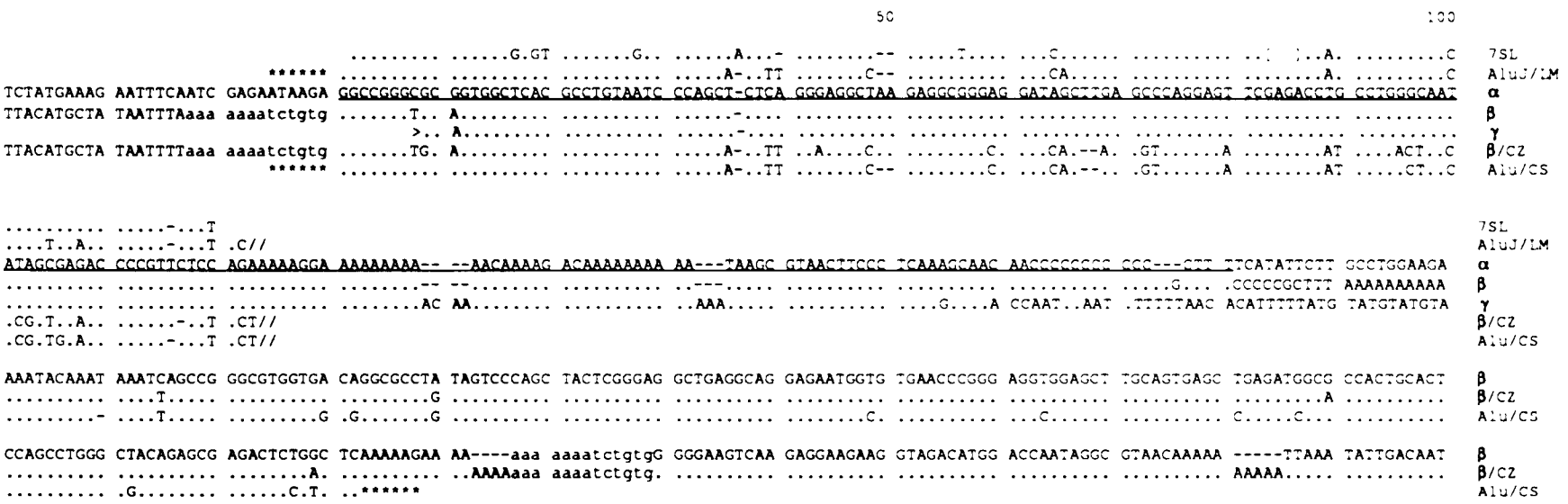
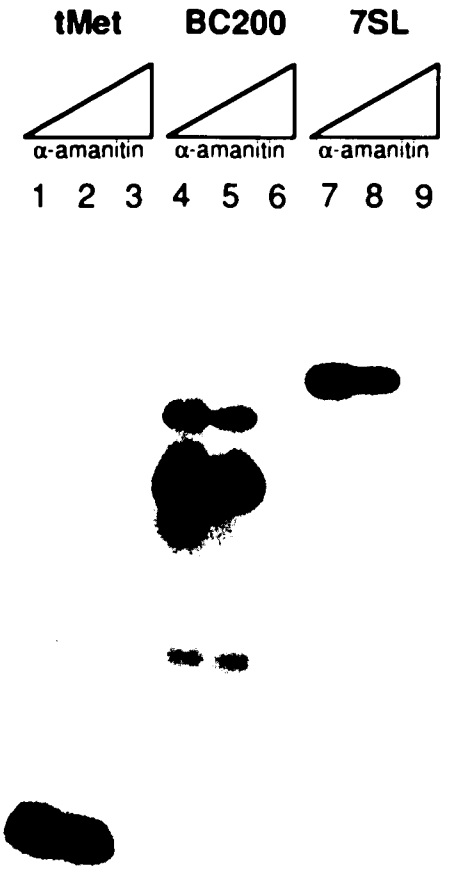

50
100
7SL
AluJ/LM
α
β
γ
β/CZ
Alu/CS

Figure 3. *In vitro* transcription of the BC200α gene. The *in vitro* transcriptional sensitivity of three gene constructs, a *X. laevis* $tRNA^{Met}$ (lanes 1-3), human BC200α (lanes 4-6), and human 7SL RNA (lanes 7-9), to α-amanitin was determined using a HeLa cell nuclear extract. For the transcription of BC200α, a plasmid containing 170 nucleotides of 5' sequence with 23 nt of 3' sequence, which effectively removed all the dimeric Alu sequences found on the 3.9 kb EcoR I construct shown in figure 1, was used. The first lane, of each triplet, contains no α-amanitin; the second lane, 2μg/ml; and the third lane, 200μg/ml. The major band seen in each triplet represents the expected product. The HeLa extracts used were a generous gift of J. Bieker and were prepared from HeLa S3 tissue culture cells[38].


tMet
BC200
7SL
α-amanitin
α-amanitin
α-amanitin
1 2 3 4 5 6 7 8 9

CHAPTER 7

CONCLUSIONS

> The history of every major Galactic Civilization tends to pass through three distinct and recognizable phases, those of Survival, Inquiry and Sophistication, otherwise known as the How, Why and Where phases. For instance, the first phase is characterized by the question How can we eat? the second by the question Why do we eat? and the third by the question Where shall we have lunch?
>
> (The Hitch Hiker's Guide to the Galaxy, Adams)

Rodent BC1 and primate BC200 RNAs are neural specific, nontranslatable, small cytoplasmic RNAs. Both molecules represent the first examples of neural specific RNA polymerase III transcribed products in their respective species. Given the variety and nature of the functions performed by other small RNAs, such as 7SL RNA (Walter and Blobel, 1982), the U snRNAs (reviewed in Reddy and Busch, 1988), RNAse P RNA (reviewed in Altman et al., 1988) and TFIIIR (Young et al., 1991), distinct neural functions could be defined by BC1 and BC200 RNAs.

This work and that of others, as previously discussed, reveals that while both BC1 RNA and BC200 RNA arose via different ancestral molecules, both progenitors are intimately involved in protein translation. Structurally, BC1 and BC200 RNAs share a similar tripartite structure: a middle repetitive element comprising their 5 end, a middle A-rich region and a 3' unique sequence. Although their 3' ends share only limited sequence identity, a probe matching the BC1 RNA 3' sequence detects BC200 RNA on Northern blots of human brain RNA. Anatomically, BC1 and BC200 RNAs are expressed in analogous subsets of neurons in the central and peripheral nervous systems. The striking dendritic and somatic localization of these structures, their association with ribonucleoprotein complexes (Kobayashi, et al., 1991; J. Brosius and J. G. Cheng, personal communication), the concurrence of BC1 RNA's onset of expression and synaptogenesis and BC200 RNA's ability to bind two protein subunits of the signal recognition particle (P. Walter, UCSF, personal communication) all support the hypothesis that BC1 and BC200 RNAs are involved in postsynaptic protein biosynthesis.

The BC1 RNA gene from rat, mouse, Chinese hamster and guinea pig have been identified and sequenced. While a cDNA copy of the BC1 RNA gene of Syrian hamster has been obtained (S. Friedman, personal communciation), the gene has not been identified. However, a BC1 RNA Syrian hamster pseudogene has been identified and sequenced. Alignment of the sequences reveals that the coding regions and certain putative regulatory elements of these genes are conserved; in marked contrast to the flanking regions which have been subject to neutral drift.

Because guinea pig represents the most evolutionary distant rodent in our sampling, and they have recently been proposed to define a

new mammalian order (Graur et al., 1991) Northern blot analyses of its brain and peripheral tissue RNA was performed. As for the myomorph rodents, guinea pig BC1 RNA is also expressed in brain but not in any of the non-neural tissues sampled. These results provide evidence that: 1. guinea pigs are rodents, not representatives of a new mammalian order; and 2. evolutionary pressure has been exerted to conserve the BC1 RNA genes and their neural-specific expression for at least 55 million years, implying that BC1 RNA is functional.

The human BC200 RNA gene is the first identified transcriptionally and transpositionally active Alu element. Structural constraints found in the RNA coding sequence suggest that the BC200 RNA gene is functional. The analysis of the full-length sequence, just as prior analysis of a partial cDNA clone sequence (Labuda and Striker, 1989; Jurka and Milosavljevic, 1991), has revealed that BC200 RNA has maintained all the CG doublets diagnostic of its progenitor Alu J sequence (described in Chapter 6). These doublets have been conserved despite the high transition rate of this dinucleotide pair and the low number present in Alu sequences. Therefore, constraints have been maintained on the BC200 RNA sequence since its estimated generation 55 million years ago.

The BC200 RNA gene has been transcriptionally active for at least 20 million years (Watson and Sutcliffe, 1987; and work presented herein). This stands in marked contrast to the *in vivo* transcriptional silence of the 500,000 members of the Alu family and the several hundred 7SL RNA pseudogenes (Paulson and Schmid, 1986; Ullu and Weiner, 1984a). The persistence of the BC200 RNA gene's transcriptional activity,

especially when viewed against this background, further suggests that evolutionary pressure also exists to maintain BC200 RNA.

This study has also addressed the transcriptional regulation of these two molecules. BC200 RNA, as determined by α amanitin sensitivity, was shown to be transcribed by RNA polymerase III. Analysis of the BC200 gene's 5' flanking region reveals the presence of a TATA box: a classic pol II promoter motif which can also regulate the transcription of some pol III genes (Geiduschek and Tocchini-Valentini, 1988; Murphy et al., 1989). Removal of the BC200 RNA gene's TATA element lowers the level of transcription. Therefore, the possibility exists that a pol II promoter is involved in regulating the pol III directed transcription of BC200 RNA. These results are also significant with regard to the generation and amplification of repetitive sequences, as discussed in Chapter 6).

To evaluate the regulatory mechanisms involved in BC1 RNA transcription, a novel homologous whole cell *in vitro* extract, prepared from rat brain tissue was developed. These tissue extracts allowed the *in vitro* transcription analysis of BC1 RNA to be performed. Cultured cells were not a potential source of extracts because BC1 RNA transcription is deregulated in these cells.

From the analysis of the BC1 RNA gene from different rodents the presence of sequences resembling RNA polymerase II (pol II) promoters were revealed. These included two octamer binding sites, a proximal sequence element and a TATA box. By using the tissue derived extracts the BC1 RNA intragenic sequences as well as its upstream region, containing the TATA, PSE and octamer binding sites, are shown to be both functional and necessary for transcription. By utilizing

mutagenized and chimaeric constructs we identified a variant B box intragenic promoter. This promoter may direct the binding of a factor which is able to mediate "communication" between pol II and III promoters. This suggests a mechanism by which the BC1 RNA gene and possibly other pol III transcribed genes can be differentially regulated. Studies of the BC1 transgenic mice generated will further our understanding of the role of *cis*-acting sequences in directing the developmental, cell type and tissue-specific expression of this RNA. Furthermore, the combination of these *in vitro* and *in vivo* techniques should reveal the underlying "network of interactions" (Sharp, 1992) present between the different polymerases.

Both BC1 RNA and BC200 RNA have most likely arisen via the process of retroposition. The result of studies presented herein strongly indicates that these two RNAs are both functional and may have even evolved convergently to fulfill analogous roles in the rodent and primate species, respectively. If functional, these two RNA species, related members of repetitive DNA families, would have important evolutionary consequences. Middle repetitive elements have been characterized as "selfish DNA"; capable of spreading throughout the genome, making no direct contribution to the fitness of the cell or organism, and whose only "function" seems to be self survival (Dawkins, 1976; Doolittle and Sapienza, 1980; Orgel and Crick, 1980). Unfortunately, the use of the anthropomorphic term belies the issue of biological consequences. The most obvious, are unfortunately detrimental, and lie in the etiology of certain genetic abnormalities such as familial hypercholesteremia, hemophilia, and gyrate atrophy (Lehrman et al., 1985; Vidaud et al., 1989; Mitchell et al., 1991). It has often been theorized that the distinction

between functional and "selfish" may be merely a matter of time (Jacob, 1982; Gould and Vrba, 1982; Rogers, 1985b). BC1 RNA and BC200 RNA may prove that individual members of what have been termed "selfish DNA" families, can acquire useful roles.

REFERENCES

Adams, D. (1979). The Hitch Hiker's Guide to the Galaxy. London: Pan Books.

Adams, T. E., Alpert, S., and Hanahan, D. (1987). Non-tolerance to a transgenic self antigen expressed in pancreatic β cells. Nature, 325, 223-228.

Altman, S., Gold, H. A. and Bartiewicz, M. (1988). In Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles, M. L. Birnstiel, ed. (Berlin: Springer-Verlag), pp 183-195.

Banker, G and Steward, O. (1992). Trends Neuro. In press.

Bantle, J. A., and Hahn, W. E. (1976). Complexity and characterization of polyadenylated RNA in the mouse brain. Cell, 8, 139-150.

Bell, S. P., Jantzen, H., and Tjian, R. (1990). Assembly of alternative multiprotein complexes directs rRNA promoter selectivity. Genes Dev., 4, 943-954.

Bell, S. P., Learned, R. M., Jantzen, H., and Tjian, R. (1988). Functional cooperativity between transcription factors UBF1 and SL1 mediates human ribosomal RNA synthesis. Science, 241, 1192-1197.

Bell, S. P., Pikaard, C. S., Reeder, R. H., and Tjian, R. (1989). Molecular mechanisms governing species-specific transcription of ribosomal RNA. Cell, 59, 489-497.

Brinster, R. L., Ritchie, K. A., Hammer, R. E., O'Brien, R. L., Arp, B., and Storb, U. (1983). Expression of a microinjected immunoglobulin gene in the spleen of transgenic mice. Nature, 306, 332-336.

Britten, R. J., and Davidson, E. H. (1969). Gene regulation for higher cells: a theory. Science, 165, 349-357.

Britten, R. J., and Davidson, E. H. (1971). Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. Q. Rev. Biol., 46, 111-138.

Bruckenstein, D. A., Lein, P. J., Higgins, D., and Fremeau Jr., R. T. (1991). Distinct spatial localization of specific mRNAs in cultured sympathetic neurons. Neuron, 5, 809-819.

Buratowski, S., Hahn, S., Guarente, I., and Sharp, P. A. (1989). Five intermediate complexes in transcription initiation by RNA polymerase II. Cell, 56, 549-561.

Burgin, K. E., Waxham, M. N., Rickling, S., Westgate, S. A., Mobley, W. C., and Kelly, P. T. (1990). In situ hybridization histochemistry of Ca^{+2}/calmodulin-dependent protein kinase in developing rat brain. J. Neurosci., 10, 1788-1798.

Carey, M. (1991). Mechanistic advances in eukaryotic gene activation. Curr. Opin. Cell Biol., 3, 452-460.

Chikaraishi, D. M. (1979). Complexity of cytoplasmic polyadenylated and nonpolyadenylated rat brain ribonucleic acids. Biochem., 18, 3249-3256.

Chikaraishi, D. M. (1986). The ID, brain identifier, model of neuronal gene expression: a re-evaluation. Trends Neurosci., 543-546.

Ciliberto, G., Castagnoli, L., and Cortese, R. (1983). Transcription by RNA polymerase III. Curr. Topics Dev. Biol., 18, 59-88.

Comai, L., Tanese, N., and Tjian, R. (1992). The TATA-binding protein and associated factors are integral components of the RNA polymerase I transcription factor, SL1. Cell, 68, 965-976.

Costantini, F., and Lacy, E. (1981). Introduction of a rabbit b-globin gene into the mouse germline. Nature, 294, 92-94.

Crichton, M. (1990). Jurassic Park. New York: Ballantine Books.

Daniels, G. R., and Deininger, P. L. (1985). Repeat sequence families derived from mammalian tRNA genes. Nature, 317, 819-822.

Dante (First published by Oxford University Press,1939). The Divine Comedy 1: Inferno (Sinclair, J. D., Trans.). New York: Oxford University Press.

Darnell, J. E. (1982). Variety in the level of gene control in eukaryotic cells. Nature, 297, 365-371.

Davidson, E. H., and Britten, R. J. (1979). Regulation of gene expression: Possible role of repetitive sequences. Science, 204, 1052-1059.

Davis, H. P., and Squire, L. R. (1984). Protein synthesis and memory. Psychol. Bull., 96, 518-559.

Davis, L., Banker, G., and Steward, O. (1987). Selective dendritic transport of RNA in hippocampal neurons in culture. Nature, 330, 474-479.

Davis, L., Burger, B., Banker, G., and Steward, O. (1990). Dendritic transport: quantitative analysis of the time course of somatodendritic transport of recently synthesized RNA. J. Neurosci., 10, 3056-3068.

Dawkins, R. (1976). The Selfish Gene. Oxford: Oxford University Press.

Deininger, P. L. (1989). SINEs: short interspersed repeated DNA elements in higher eukaryotes. In D. E. Berg and M. M. Howe (Eds.), Mobile DNA (pp. 619-636). Washington D.C.: American Society for Microbiology.

Deininger, P. L., Batzer, M. A., Hutchison III, C. A., and Edgell, M. H. (1992). Master genes in mammalian repetitive DNA amplification. Trends Genet., in press.

Dombroski, B. A., Mathias, S. L., Nanthakumar, E., Scott, A. F., and Kazazian Jr., H. H. (1991). Isolation of an active human transposable element. Science, 254, 1805-1808.

Donovan, D. M., Takemura, M., O'Hara, B. F., Brannock, M. T., and Uhl, G. R. (1992). Preproenkephalin promoter "cassette" confers brain expression and synaptic regulation in transgenic mice. Proc. Natl. Acad. Sci. USA, 89, 2345-2349.

Doolittle, W. F., and Sapienza, C. (1980). Selfish genes, the phenotype paradigm and genome evolution. Nature, 284, 601-603.

Folk, W. R. (1988). Changing directions in pol III transcription. Genes Dev., 2, 373-375.

Gabrielsen, O. S., and Sentenac, A. (1991). RNA polymerase III (C) and its transcription factors. Trends in Biochem., 16, 412-416.

Garner, C. C., Tucker, R. P., and Matus, A. (1988). Nature, 336, 674-677.

Geiduschek, E. P., and Tocchini-Valentini, G. P. (1988). Transcription by RNA polymerase III. Ann. Rev. Biochem., 57, 873-914.

Gordon, J., Scangos, G., Plotkin, D., Barbosa, J., and Ruddle, F. (1980). Genetic transformation of mouse embryos by microinjection of purified DNA. Proc. Natl. Acad. Sci. USA, 77, 7380-7384.

Gould, S. J. and Vrba, E. S. (1982). Exaptation - a missing term in the science of form. Paleo., 8, 4-15.

Graur, D., Hide, W. A., Li, W.-H. (1991). Is the guinea pig a rodent? Nature, 351, 649-652.

Grosveld, F., van Assendelft, G. B., Grenves, D. R., and Kollias, G. (1987). Position-independent, high level expression of the human b-globin gene in transgenic mice. Cell, 51, 975-985.

Guarente, L. (1988). UASs and enhancers: common mechanism of transcriptional activation in yeast and mammals. Cell, 52, 303-305.

Hanahan, D. (1988). Dissecting multistep tumorigenesis in transgenic mice. Ann. Rev. Genet., 22, 479-519.

Harbers, K., Jahner, D., and Jaenisch, R. (1981). Microinjection of cloned retroviral genomes into mouse zygotes: integration and expression in the animal. Nature, 293, 540-547.

Hastie, N. D., and Bishop, J. O. (1976). The expression of three abundance classes of messenger RNA in mouse tissues. Cell, 9, 761-774.

Hipskind, R. A., and Clarkson, S. G. (1983). 5' flanking sequences that inhibit *in vitro* transcription of a *Xenopus laevis* tRNA gene. Cell, 34, 881-890.

Howard, B. H., and Sakamoto, K. (1990). Alu interspersed repeats: selfish DNA or a functional gene family? New Biologist, 2, 759-770.

Jacob, F. (1982). Molecular tinkering in evolution. in Evolution from Molecules to Men (Cambridge University Press) pp 131-144.

Jagadeeswaran, P., Forget, B. G., and Weissman, S. M. (1981). Short interspersed repetitive DNA elements in eucaryotes: transposable DNA elements generated by reverse transcription of RNA pol III transcripts? Cell, 26, 141-142.

Jelinek, W. R., and Schmid, C. W. (1982). Repetitive sequences in eukaryotic DNA and their expression. Annu. Rev. Biochem., 51, 813-844.

Johnson, P. F., and McKnight, S. L. (1989). Eukaryotic transcriptional regulatory proteins. Annu. Rev. Biochem., 58, 799-839.

Jurka, J. and Milosavljevic, A. (1991). Reconstruction and analysis of human Alu genes. J. Mol. Evol., 32, 105-121.

Jurka, J. and Zuckerkandl, E. (1991). Free left arms as precursor molecules in the evolution of Alu sequences. J. Mol. Evol., 33, 49-56.

Kadonaga, J. T. (1990). Gene transcription: basal and regulated transcription by RNA polymerase II. Curr. Opin. Cell Biol., 2, 496-501.

Karin, M. (1990). Too many transcription factors: positive and negative interactions. New Biol., 2, 126-131.

Kedes, L. H. (1979). Histone genes and histone messengers. Ann. Rev. Biochem., 48, 837-870.

Kleiman, R., Banker, G., and Steward, O. (1991). Differential subcellular localization of particular mRNAs in hippocampal neurons in culture. Neuron, 5, 821-830.

Kleinert, H., and Benecke, B.-J. (1988). Transcription of human 7SK DNA *in vitro* and *in vivo* is exclusively controlled by an upstream promoter. Nucl. Acids Res., 16, 1319-1331.

Kobayashi, S., Goto, S., Anzai, K. (1991). Brain-specific small RNA transcript of the identifier sequences is present as a 10 S ribonucleoprotein particle. Jour. Biol. Chem., 266, 4726-4730.

Kunkel, G. R. (1991). RNA polymerase III transcription of genes that lack internal control regions. Bioch. Biophys. Acta, 1088, 1-9.

Labuda, D., and Striker, G. (1989). Sequence conservation in Alu evolution. Nucl. Acids Res., 17, 2477-2491.

Lassar, A. B., Martin, P. L., and Roeder, R. G. (1983). Transcription of class III genes: formation of preinitiation complexes. Science, 222, 740-748.

Learned, R. M., Cordes, S., and Tjian, R. (1985). Purification and characterization of a transcription factor that confers promoter specificity to human RNA polymerase I. Mol. Cell. Biol., 5, 1358-1369.

Learned, R. M., Learned, T. K., Haltiner, M. M., and Tjian, R. T. (1986). Human rRNA transcription is modulated by the coordinate binding of two factors to an upstream control element. Cell, 45, 847-857.

Levine, M., and Manley, J. L. (1989). Transcriptional repression of eukaryotic promoters. Cell, 59, 405-408.

Lobo, S. M., and Hernandez, N. (1989). A 7 bp mutation converts a human RNA polymerase II snRNA promoter into an RNA polymerase III promoter. Cell, 58, 55-67.

Lobo, S. M., Lister, J., Sullivan, M. L., and Hernandez, N. (1991). The cloned RNA polymerase II transcription factor IID selects RNA polymerase III to transcribe the human U6 gene in vitro. Genes Devel., 5, 1477-1489.

Lone, Y.-C., Simon, M.-P., Kahn, A., and Marie, J. (1986). Sequences complementary to the brain-specific "identifier" sequences exist in L-type pyruvate kinase mRNA (a liver-specific messenger) and in transcripts especially abundant in muscle. Jour. Biol. Chem., 261, 1499-1502.

Long, E. O., and Dawid, I. B. (1980). Repeated genes in eukaryotes. Annu. Rev. Biochem., 49, 727-764.

Lucretius (First published by Penguin Books,1951). On the Nature of the Universe (Latham, R. E., Trans.). New York: Penguin Books.

Lund, E., Dahlberg, J. E., and Forbes, D. J. (1984). The two embryonic U1 small nuclear RNAs of *Xenopus laevis* are encoded by a major family of tandemly repeated genes. Mol. Cell. Biol, 4, 2580-2586.

Maniatis, T., Goodbourn, S., and Fischer, J. A. (1987). Regulation of inducible and tissue-specific gene expression. Science, 236, 1237-1245.

Margottin, F., Dujardin, G., Gerard, M., Egly, J.-M., Huet, J., and Sentenac, A. (1991). Participation of the TATA factor in transcription of the yeast U6 gene by RNA polymerase C. Science, 251, 424-426.

Mathias, S. L., Scott, A. F., Kazazian Jr., H. H.,Boeke, J. D., and Gabriel, A. (1991). Reverse transcriptase encoded by a human transposable element. Science, 254, 1808-1810.

Mattaj, I. W., Dathan, N. A., Parry, H. D., Carbon, P., and Krol, A. (1988). Changing the RNA polymerase specificity of U snRNA gene promoters. Cell, 55, 435-442.

McKinnon, R. D., Shinnick, T. M., and Sutcliffe, J. G. (1986). The neuronal identifier element is a *cis*-acting positive regulator of gene expression. Proc. Natl. Acad. Sci. USA, 83, 3751-3755.

Milner, R. J., Bloom, F. E., Lai, C., Lerner, R. A., and Sutcliffe, J. G. (1984). Brain-specific genes have identifier sequences in their introns. Proc. Natl. Acad. Sci. USA, 81, 713-717.

Milner, R. J., and Sutcliffe, J. G. (1983). Gene expression in rat brain. Nucl. Acids Res., 11, 5497-5519.

Mitchell, P. J., and Tjian, R. (1989). Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science, 245, 371-378.

Montarolo, P. G., Goelet, P., Castelluci, V. F., Morgan, J., Kandel, E. R., and Schacher, S. (1986). A critical period for macromolecular synthesis in long-term heterosynaptic facilitation in Aplysia. Science, 234, 1249-1254.

Murphy, S., Altruda, F., Ullu, E., Tripodi, M., Silengo, L., and Melli, M. (1984). DNA sequences complementary to human 7SK RNA show structural similarities to the short mobile elements of the mammalian genome. J. Mol. Biol., 177, 575-590.

Murphy, S., DiLiegro, C., and Melli, M. (1987). The in vitro transcription of the 7SK RNA gene by RNA polymerase III is dependent only on the presence of an upstream promoter. Cell, 51, 81-87.

Murphy, S., Moorefield, B., and Pieler, T. (1989). Common mechanisms of promoter recognition by RNA polymerase II and III. Trends Genet., 5, 122-126.

Orgel, L. E., and Crick, F. H. C. (1980). Selfish DNA: the ultimate parasite. Nature, 284, 604-607.

Owens, G. P., Chaudhari, N., and Hahn, W. E. (1985). Brain "identifier sequence" is not restricted to brain: similar abundance in nuclear RNA of other organs. Science, 229, 1263-1268.

Palmer, J. M., and Folk, W. R. (1990). Unraveling the complexities of transcription by RNA polymerase III. Trends Biochem., 15, 300-304.

Palmiter, R. D., and Brinster, R. L. (1986). Germ-line transformation of mice. Annu. Rev. Genet., 20, 465-499.

Parvin, J. D., Timmers, H. T. M., and Sharp, P. A. (1992). Promoter specificity of basal transcription factors. Cell, 68, 1135-1144.

Paulson, K. E., and Schmid, C. W. (1986). Transcriptional inactivity of Alu repeats in HeLa cells. Nucl. Acids Res., 14, 6145-6158.

Piatagorsky, J., and Wistow, G. (1991). The recruitment of crystallins: new functions precede gene duplication. Science, 252, 1078-1079.

Pikaard, C. S., McStay, B., Schultz, M. C., Bell, S. P., and Reeder, R. H. (1989). The Xenopus ribosomal gene enhancers bind an essential polymerase I transcription factor, xUBF. Genes Dev., 3, 1779-1788.

Quentin, Y. (1992). Fusion of a free left Alu monomer and a free right Alu monomer at the origin of the Alu family in the primate genomes. Nucl. Acids Res., 20, 487-493.

Rao, A., and Steward, O. (1991). Evidence that protein constituents of postsynaptic membrane specializations are locally synthesized: analysis of proteins synthesized within synaptosomes. J. Neurosci., 11, 2881-2895.

Reddy, R., and Busch, H. (1988). Small nuclear RNAs: RNA sequences, structure, and modifications. In Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles, M. L. Birnstiel, ed. (Berlin: Springer-Verlag), pp 1-37.

Reddy, R., Li, W.-Y., Henning, D. Y., Choi, C., Nohga, K., and Busch, H. (1981). Characterization and subcellular localization of 7-8 S RNAs of Novikoff Hepatoma. J. Biol. Chem., 256, 8452-8457.

Reeder, R. H. (1990). rRNA synthesis in the nucleus. Trends Genet., 6, 390-395.

Renkawitz, R. (1990). Transcriptional repression in eukaryotes. Trends Genet., 6, 192-196.

Roeder, R. G. (1991). The complexities of eukaryotic transcription initiation: regulation of preinitiation complex assembly. Trends. Biochem., 16, 402-408.

Rogers, J. (1983). Retroposons defined. Nature, 301, 460.

Rogers, J. (1985a). The origin and evolution of retroposons. Int. Rev. Cytol., 93, 187-279.

Rogers, J. (1985b). Origins of repeated DNA. Nature, 317, 765-766.

Rosenfeld, M. G., Crenshaw, E. B., and Lira, S. A. (1988). Transgenic mice: applications to the study of the nervous system. Annu. Rev. Neurosci., 11, 353-372.

Sakamoto, K., and Okada, N. (1985). 5-methylcytidylic modification of in vitro transcript from the rat identifier sequence; evidence that the transcript forms a tRNA-like structure. Nucl. Acids Res., 13, 7195-7206.

Sawadogo, M., and Roeder, R. (1985). Interaction of a gene-specific transcription factor with the adenovirus major late promoter upstream of the TATA box region. Cell, 43, 165-175.

Sawadogo, M., and Sentenac, A. (1990). RNA Polymerase B (II) and general transcription factors. Annu. Rev. Biochem., 59, 711-754.

Sharp, P. A. (1983). Conversion of RNA to DNA in mammals: Alu-like elements and pseudogenes. Nature, 301, 471-472.

Sharp, P. A. (1992). TATA-binding protein is a classless factor. Cell, 68, 819-821.

Simmen, K. A., Bernues, J., Parry, H. D., Stunnenberg, H. G., Berkenstam, A., Cavallini, B., Egly, J.-M., and Mattaj, I. W. (1991). TFIID is required for in vitro transcription of the human U6 gene by RNA polymerase III. EMBO J., 10, 1853-1862.

Singer, M. F. (1982). SINEs and LINEs: highly repeated short and long interspersed sequences in mammalian genomes. Cell, 28, 433-434.

Smith, L.M., Sanders, J. Z., Kaiser, R. J., Hughes, P., Dodd, C., Connell, C. R., Heiner, C. Kent, S. B. H., and Hood, L. E. (1986). Fluorescence detection in automated DNA sequence analysis. Nature, 321, 674-679.

Sollner-Webb, B. (1988). Surprises in polymerase III transcription. Cell, 52, 153-154.

Sollner-Webb, B., and Mougey, E. B. (1991). News from the nucleolus: rRNA gene expression. Trends Biochem., 16, 58-62.

Sollner-Webb, B., and Tower, J. (1986). Transcription of cloned eukaryotic ribosomal RNA genes. Annu. Rev. Biochem., 55, 801-830.

Steward, O. (1983). Alterations in polyribosomes associated with dendritic spines during the reinnervation of the dentate gyrus of the adult rat. J. Neurosci., 3, 177-188.

Steward, O., Davis, L., Dotti, C., Phillips, L. L., Rao, A., and Banker, G. (1988). Protein synthesis and processing in cytoplasmic microdomains beneath postsynaptic sites on CNS neurons: a mechanism for establishing and maintaining a mosaic postsynaptic receptive surface. Mol. Neurobiol., 2, 227-261.

Steward, O., and Falk, P. M. (1986). Protein synthetic machinery at postsynaptic sites during synaptogenesis: a quantitative study of

the association between polyribosomes and developing synapses. J. Neurosci., 6, 412-423.

Steward, O., and Levy, W. B. (1982). Preferential localization of polyribosomes under the base of dendritic spines in granule cells of the dentate gyrus. J. Neurosci., 2, 284-291.

Steward, O., and Reeves, T. M. (1988). Protein-synthetic machinery beneath postsynaptic sites on CNS neurons: association between polyribosomes and other organelles at the synaptic site. J. Neurosci., 8, 176-184.

Sutcliffe, J. G., Milner, R. J., Bloom, F. E., and Lerner, R. A. (1982). Common 82-nucleotide sequence unique to brain RNA. Proc. Natl. Acad. Sci. USA, 79, 4942-4946.

Sutcliffe, J. G., Milner, R. J., Gottesfeld, J. M., and Lerner, R. A. (1984a). Identifier sequences are transcribed specifically in brain. Nature, 308, 237-241.

Sutcliffe, J. G., Milner, R. J., Gottesfeld, J. M., and Reynolds, W. (1984b). Control of neuronal gene expression. Science, 225, 1308-1315.

Tiedge, H., Fremeau Jr., R. T., Weinstock, P. H., Arancio, O., and Brosius, J. (1991). Dendritic location of neural BC1 RNA. Proc. Natl. Acad. Sci. USA, 88, 2093-2097.

Tiedge, H., Chen, W., Brosius, J. (1992). Primary structure and neural-specific expression of human BC200 RNA. Submitted.

Torre, E. R., and Steward, O. (1991). Demonstration of local protein synthesis within dendrites using a new cell culture system which permits the isolation of living axons and dendrites from their cell bodies. J. Neurosci.

Ullu, E., and Tschudi, C. (1984). Alu sequences are processed 7SL RNA genes. Nature, 312, 171-172.

Ullu, E., and Weiner, A. M. (1984a). Human genes and pseudogenes for the 7SL RNA component of signal recognition particle. EMBO J., 3, 3303-3310.

Ullu, E., and Weiner, A. M. (1984b). Upstream sequences modulate the internal promoter of the human 7SL RNA gene. Nature, 318, 371-374.

Van Arsdell, S. W., Denison, R. A., Bernstein, L. B., and Weiner, A. M. (1981). Direct repeats flank three small nuclear RNA pseudogenes in the human genome. Cell, 26, 11-17.

Van Arsdell, S. W., and Weiner, A. M. (1984). Human genes for U2 small nuclear RNA are tandemly repeated. Mol. Cell. Biol., 4, 492-499.

Van Dyke, M. W., Roeder, R., and Sawadogo, M. (1988). Physical analysis of transcription preinitiation complex assembly on a class II gene promoter. Science, 241, 1335-1338.

Wagner, E. F., Stewart, T. A., and Mintz, B. (1981). The human β-globin gene and a functional viral thymidine kinase gene in developing mice. Proc. Natl. Acad. Sci. USA, 78, 5016-5020.

Waibel, F., and Filipowicz, W. (1990). RNA-polymerase specificity of transcription of Arabidopsis U snRNA genes determined by promoter element spacing. Nature, 346, 199-202.

Walter, P., and Blobel, G. (1982). Signal recognition particle contains a 7S RNA essential for protein translocation across the endoplasmic reticulum. Nature, 299, 691-698.

Wawrousek, E. F., Chepelinsky, A. B., McDermott, J. B., and Piatigorsky, J. (1990). Dev. Biol., 137, 8-76.

Weiner, A. M., Deininger, P. L., and Efstratiadis, A. (1986). Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. Annu. Rev. Biochem., 55, 631-661.

Wolffe, A. P. (1991). RNA polymerase III transcription. Curr. Opin. Cell Biol., 3, 461-466.

Young, L. S., Dunstan, H. M., Witte, P. R., Smith, T. P., Ottonello, S., Sprague, K. U. (1991). A class III transcription factor composed of RNA. Science, 252, 542-546.

Zieve, G., and Penman, S. (1976). Small RNA species of the HeLa cell: metabolism and subcellular localization. Cell, 8, 19-31.

APPENDIX

The following pages contain the sequences of the rodent BC1 RNA genes and a pseudogene, human BC200 RNA gene and pseudogenes and the sequence of the corresponding locus of the BC200β gene locus from chimpanzee. The coding region of each sequence has been double underlined and the presence of B1 repetitive elements in the rodent sequences and Alu elements in the human sequences has been noted by dotted lines. The orientation of these repetitive elements, with respect to their internal promoter sequences, is noted in the text.

RAT BC1 RNA GENE Length:2233

```
   1  CCCGGGCATA CAGTCCTGTA ATCCTAGCCC AGGTTGGGTG CAAGATGCGC TCAGCCCTTT CCACTCCCTG AGGGTAGCAG GCATATGAGG TTTTCAACTC
 101  TATTTGGTTG ACCACTTATC ATCTGTCCAT TCAGAAGGTG GAGCAGGAGG GTCAGGAGTT CAAGACTAGC ACATAGGGAT GCTGTTTCAA ACTCACCCTA
 201  CCCTCACCCA CTGAAAAGTG CCTCTTCTAC TGAGGAATCC TCCAGCTTCT TTCCTTTCCT TTCTTTTTTC TTTTTCTTT TTTCTTTTTT TTTTTTTTG
 301  GATTTTGTTT GTTTGTTTCT GGTTTTAAGT TCTGAAGACA GGGTTTCTCT GTGTAGCCCT GGCTGTCCCG GAACTCTGTT TATAAACCAG GCTGGCCTTG
 401  AACTCTCAGA GATCCACCTG CCTCAACTTC CCAAGTACTG GGATTAAAGG AAAGTCTTTG CCACCAGACC TGGCCTCCTG CCTCCTTTTT CTTTTTTATG
 501  GTTAGCTCTT TACTTGGTGT AAGTGAACAG CAGCGTGAGC AGCATGCGCA GCGTGAGCAG CGTGAGCAGG GTGTGAGGGC CTCACAGGAC TTTGTTTTCC
 601  TGACAAACCA ATACTTTCCT ATCATGTGGA GTGCTTACTG CTCTATCCTC CCCCGAGCTT GTGATTCTAT TTTTATCATG TGAAGTTTTA CTCCTAGGAG
 701  ACTTCACCTT TGTCCTTTCT AGTCTGTTCT CCAGTCAACA GGCAACCATA CTGGTTGTGA GACTCAAACT TCTCTCTGCA CAGGAGCTCT CATTTTCAAA
 801  GAGCACCATG TTAGAGTTAA GGATTCATTT GTATATATGA AATTCAGATT GGAGGTAGAA TCCCTGGAGT GTGTGTGTAT GTGTGTGTGT GTGTTCACTT
 901  AGCCCTCCAA TCAATTATAT TGTGTCTCTG GTCTCTTTCA TACAGAAGCT GATCAACTTG TTAGCAATTG TTGTCTAAGG AATTTACTGT ACTTTGGAAT
1001  ACTTTTTTAA ACTATATTTT GGAAGGTATC TCTGATGTAA ATGAGAAAGC TAACTTCAGT TTCTGCTTTT TGGAATGTGG GTGCCTATGG GATTGGAGCG
1101  CTTGGACAAA GTGGCTCCTC CTGCCGGCCA GCCCTTCGGT ACCATCTGAT ACTTGACTGT GTATGAAATA CTTCAGTTTT GCTTTAAATA CTGGCAGCAA
1201  GAGCTAACGT TCGGGGGTTGG GGATTTAGCT CAGTGGTAGA GCGCTTGCCT AGCAAGCGCA AGGCCCTGGG TTCGGTCCTC AGCTCCGAAA AAAAAAAAAA
1301  AAAAAAAAAG ACAAAATAAC AAAAAGACCA AAAAAAAACA AGGTAACTGG CACACACAAC CTTTTTCATT TTCAAAGACC CCCAAGGGCA TTTTCATGGC
1401  TGGGGGCAAT ATTTATTATC TGCCCTTTAG GATATCGCAG TATCAAATCG TTTCCAGTCC GCAGGGAAGC TCCCGTGCAG TGTACAGATG CGATGATTGA
1501  CAGCCCGGTG ACAGTGGAAG ATGCCGGATG CTCCTCCCTC CTTGAAGCCT GCCCCTAGGG CCCTTCCTGG ATTTCCTGGG AACGCCCCTG TGCCAAATCT
1601  TTGATAACTT TTCTTATTTC TTTAAGACCC ATGTTTGTGA ATTAGGTTCC ATCTGTTTGG ATTTTGGAAA TAAAAGCAGG ATTCATCTGG GCCCCGCGCG
1701  CGACCCAGAG GACACGCCCC AAGGGATTGG CCCGACCCCA CCCACCACTG AACCAGCCGT GCCAGGGAGG AGGAGCAAGG CTTCACTCGG GCTAGCCCCG
1801  CCCACATATT ACCTACGCAT GCGTAGAGAG GTGGGCTTTT TTTTTTTTTT TTTTTGGCTT CTTTAGCTTG CCGTCCTTAC CCGCCCTGCT CCGCGCGGCA
```

1901 TAAACGGAAG TGGCTTCTTC AATCTGGGCA TTTGTGCTCG TGTCTCGCCG GCTGGCGATT GCTACACAAA ATGACAGGAG TGCGAGGACC CGGGGAGAGC

2001 CGAGACACCG CCATGGCCAT GGAGCAGGAC ATATTTGACG CCGTCGTGAT GGCGGATGAG AGGTGGGCAG AGAGATCCGG CTCCGCCCCG CCCCGGCAGC

2101 TGCTGTGCCG TCCTCGGGGA TACCCGCCTT TGCCATCTGC GAGCTGCCAC TTTTTCTTGT CCTTTAATTG CCGCACATAC GGCGGTATTC ATGAATCTGC

2201 TGATGGATAT GTCGGACAGG CCCCTTTGGA TCC

B

MOUSE BC1 RNA GENE Length:1695

```
   1 GAATTCATTC ATTCATTCAT CTATTTCTTT TTTTTTTTGT TTTTTTGTTT TTTTGTTTTT TTTCGAGACA GGGTTTCTCT GTGTAGCCCT GGCTGTCCTG
 101 GAACTCACTT TGTAGACCAG GCTGGCCTCG AACTCAGAAA TCCGCCTGCC TCTGCCTCCC AAGTGCTGGG ATTAAAGGCG TACGCCACAA CGCCCGGCTC
 201 ATTCATCTAT TTCTAAACAA TTATATTTTA TGTTTTGAGA TATGGTCTCA CTCTTCAGCC CAGACCAGTC TGGAATGCAC TCTGTAGCCC AGGTTGGCCT
 301 CAGATTCATG GCAGTTCTTT TCCCTCAGCT TCCCACATGG GGCTTACAAG TGGGGGCCAC TGTCCCTGCT TCCAGGATGA ATCTGTTTTG TACCCACATC
 401 TTTTTTTTTC TTTTCATTTT TTGAGACAGG TTTTCTCTGT GTAGCCCTGG CTGTCCTGGA ACTTACTCTG TAGACCAGGC TGGCCTCGAA CTCAGAAATC
 501 TGCCTGCCTC TGCCTCCGAA GTGCTGGGAT TAAAAGAAGA CCTTTGCCAC CGCACCTGGC CTCCTTTTTC TTTTTGTGGT TAACTTTTTA CTTGGTGTAG
 601 ACTCTTCTGA TGAGCAGTGT GAGCAGCGTG AGCAGAGTGT GAGGGACTCA CTTTATTTTC CTGAGAAACC ATACTTGCCT GTCATGGCGA GTGCTCACTG
 701 CTCTATCCTT CCTGAAGTTT GTGATTTTAT TTTTATCACG GAAGTTTTAC TCCTAGGAGA CATTCTAGGC TGTTCTCCAG GCTTATCTCT TCTTCAGTCA
 801 ACATGCCATC ATGTTCATCA CCAGACTCAA ACTTCTCTCT TCTGTCTGAC ACAGTAGTTC TCATTTTCAA AAAGCATTTT GTTATAGTTA AGGATTCATT
 901 TGCACAGATG AAATTCAGTT CAAGTGGCGT GTTTACTTAT TCAGCCCTCC ATCAATTACA GTGTTGTATC TCTGGTCTCT TTTATACAGG AGCTGATTAA
1001 CTTGTTAGCA ATTGTTGCCT AAGGAACTTA CTGTACTGTG GAATACTCTT TAAAACTATA CTTTGGAAGG TATCTCTGAT GGAAATGATG AGAAAACTAA
1101 CTTCAATTTC TGCTGTTTGA AGAGTGGGTG CCTGTGAGAT TGGAGCGCTT GGACAAAGTG GCTCCTCCTG CCGGCCAGCC CTTCGGAACC ACCTGATACT
1201 TGACTGTGTA TGAAATACTT CAGTTTTGCT TTAAATACTG GCAGCAAGTG CTAGCGTTCG GGGTTGGGGA TTTAGCTCAG TGGTAGAGCG CTTGCCTAGC
1301 AAGCGCAAGG CCCTGGGTTC GGTCCTCAGC TCTGGAAAAA AAAAAAAAAA AAAAAAGACA AAATAACAAA AAGACCAAAA AAAAACAAGG TAACTGGCAC
1401 ACACAACCTT TTTCATTTTC AAAGACCCCC CCCCAAGGGC ATTTTCATGT CTGGGGGCAA TATTTATTAT CTGCCCTTTA GGATATCGCA GTATCAAATC
1501 GTTTCCAGCC CTCAGGGAGG GTCCGGTGCA GTGTACAGAT GCGGTGATTG ACAGTCCGGT GACAGCAGAG GATGCCGGAT ACTCCTCCCT CCTTGAAGCC
1601 TGCCCCTTGG GCCCTTCCTC CATTTCCTGG GAACGCCCCT GTGCCAGATC TTTGGTAACT TTTCTCATTT CTTTAAGACC CACGTTTGTG AATTC
```

CHINESE HAMSTER BC1 RNA GENE Length:1104

```
   1  AAGCTTTTTT TTTTTTTTTT TTTTTTGTAT CTGACATGGT AGTTCTTATTAAAAAAAAAA AAAAAAACAC CTCGTTATAG TTAAGGATTT ATTTGCACAG
 101  ATGAAATTCA GACCTGGAGG CAGCTTCCGA ATCCCATGGA GTTTTGTTGC CATTGTTCAC TTGTTCAGCC CTCTGACCCA GAATAATGGT GTATCTCTGG
 201  TCTCCTTCAC AGAGAGACTA TTTAACTTGT TAGTAATTGT GGTCTAAGAA ACTTACTGTG GCTCCCACAA GGCCTCATTT CCATCGGCAG TCTGGTTCAG
 301  GCTATGGCTG GTAGATCCTG TCCATTGCTA TTTTAGATGT CTATTTGGCT GTCATGGTGG GTTCCACATC TGTGATAGTG GCACCCTTGT CCCCTTGGTG
 401  CATAATCTTG CATGATAAGA ATTTTCTTGT CTTTCCCTGA TTTATTTACA CGCGAGAGCT CCTGGAATCT CAGATCTCCT TCTTGACAGT CTGAAATACC
 501  CTTTAATGTT CTACTTTGGA AGGTATCTCT GATGGAAATG AAGAGAGAAT TAACTTCAAT TCCTGCTTTT TGGAGAGTGG GTGCCTATGG GATGGGAGCG
 601  CCTGGACAGA GTGGCTCCTC CTGCTGGCAA GCCCTTCGGT ACCACCTGAT ACTTGACTGT GTATGAAATA CTTCAGTTTT GCTTTAAATA CTGGCAGCAA
 701  GAGCTAGCTT TCGGGGTTGG GGATTTAGCT CAGTGGTAGA GCGCTTGCCT AGCAAGCGCA AGGCCCTGGG TTCGGTCCTC AGCCCTGGAA AAAAAAAAAA
 801  AAAAAAAAAA AAGACAAAAT AACAAAAAAG ACCAAAAAAA AACAAGGTAA CTGGCACACA CAACCTTTTT CATTTTCAAA GGCCCCTCAA GGAATTTTCC
 901  TGGCTGGGGG AAGTATTTAT TATCTGCCCT TTAGGATGTC GCAGTATCAA ATCGGTTCCA GCTCGCAGAT AAGGCTCCTG TGTGTGCAGA TGCAGTGATT
1001  GACAGCCGGA TGGCAGCTGA GGATGCTGGA TGCTCCACCT CCGTCCGTGC GGCTTGCAGC TGGGGCCCTT TCTTCACTTC ATGGGAATGC CCCTGTGCTA
1101  GATC
```

D

GUINEA PIG BC1 RNA GENE Length:1307

```
   1  GGATCCTGGT GTATTTATCT CTAGCTTGTT TCACATGAGT CCTGTTAGAA ATCGTTACTT CTAAGGAATT TCTGACTACT GTAAAGCAAT ACCCCATCAC
 101  CTTCCTGCAC AGGGAACTGC TAATCATTGT TGGAACAGCA GAAAGCAAGC AGTTTTATCT TAGTACACAT GTGTGTTGCT ATTTGCTGGA CTCTTAGGAA
 201  TTCTGAGGCT TCAAATAATT TTTTGTGGGT GGCATGACAC TTGGAAATAG TGATACTTTT TCTCTTTCTA ATGCTTACAC TTCTTGTGTG TCATGTTACT
 301  GCATGATCTG GGTCTAGCAG AAAAATATCC GGTTAGCATA TATGTGATAA AGGCATTTTT GTTTTATTCT TCATTTGAAA GGAAGTATTT CTAGTGTTTC
 401  ACATGTATGT TTTGTGGTTG GTAGTTTCAA ATAGCTTGTT AAAAATCTTG TTCAGAAATT ATCTCTCATG AAGATGACCT AGGAGAGCTA AAACCCAATC
 501  CTGTGGAATT TTTGTTGTTG TTTTCATAGG AACGGAGGCC TCCAGTTGGA ATACCTGGGA AGGGGGAGGG GGTCAAACTT CTCTAAAGTT TCTTATGCAA
 601  CATGAATCAT CTTGTATTGG CTTTAAATTC ACAGAGCAAC AGCCAACATT CGGGGTTGGG GATTTGGCTC AGTGGTAGAA CGCTTGCCTA GCAAGCTGGA
 701  AACCCTGGGT TCGGTCCTCA GCTCCAAACC TGAAAAACAA AAAATCCCTA TAAAAAAACA CAAAATATAA AAAAGACCAA AAACAAACAA GGTAACTGCG
 801  CACACAAACA ACCTTTTGTA TTTTACCTTC CAAAGAACTT CTAACATACC AAGAGCAGAT GGCCACATTT AATATTTATT ATGTTTTGGA TGTCCCTGTG
 901  CTAATCGGCC CTTCCTCCTT GGTCACATCC AGCCGCTTGC TGCAGTCCGG CGGTGTGGCC GCGGACCCTC ACGCTGCTTG CACCGCGCCC AGCCCTCTGC
1001  ACCCCGCCGC GGGCTCTCCA GGACTGCGCG ACCCGACCCG CAGCCCCAAA TCCCACGGGC AGAGCTTCCG GGCCCGGAGC ATCCTGAGCA GCCGGTTCTT
1101  TGATAACTTT TTCATTTCAA TTCACTCACA TCGGCCTGTC TTTTCAAAAA CTAAAATCGG GGTCGGCAGC CCCCAACCTC AGACAACTGG AAAGTCTGGG
1201  CGGTGACAGA CCCACCAGTG GTGGAGCAGA GGGGGAGGAG AAGCGGAAAA TGGGCGGGGC TGCATTGAAG CCCTGACCCG CCCACTGCCG CAGTTGTGCT
1301  TGGATCC
```

SYRIAN HAMSTER BC1 RNA PSEUDOGENE Length:1398

```
   1 CTGCAGGTAC AGTTTAAGGT AGAATACTTG CCTAATATGT AAAAAGCCCT GGTTTTGCTC CCTGATGTTG TAAGAAAATG AATAAATAAA AAAGTAAATA
 101 GGCCTGAGGA GATGGCTCAT CAGATAAAGG CACTTGTTGC CTGCCCATCT TTAGGGCACA TGGTGAAAGG GGAGAACTGA CTTCTGAAAG TAGTCCTCTG
 201 GCTTCCACAT ATGTATGTGT ATTTGTTTAC ACACACCACA CATACACACA CACACACACA CACCACACAC CACATACACC ACACACACTC ACTCTAAATG
 301 AAAAAAATGT AAAATAAATT AAAGCCAGGC ATGATAACAC ACATTGAAAT TTACAGCTAT GGCCAGGCAT TGGTGGCGCA TGCCTTCAAT CCCAGCACTC
 401 AGGAGGCAGA GGCAGGTGGA TCTCTGTGAG TTCGAAACCA AGCCTAGTCT ACAGAGCGGG TTCCAGGACA GGCTCCAAAA CAATACAGAG AAACCTGTCT
 501 CAAAAAACAA AAAACAAAAA CAAAATATAG CTATGGAGTA AAACTGTCTT AAGAGGGGGT TGGCCAGGGC AGTGGTGGTT CACACCACTT TAAAATAAAA
 601 CAAAATCACC CCCCACAAAA GGAGGAGGAA GACAACTTGC TTACCCCTGA ATATTGCTAG TTTTATTATC AAATATTCTT TTTGTTTTGT TTTGTTTTTC
 701 GAGACAGGGT TTCTCTGAGT AACTGAACAT TTTCTTAAAA AAGCAGAATA CCGGGTTGGG GATTTAGCTC AGTGGTAGAG CGCTTGCCTA GCAAGCGCAA
 801 GGCCCTGGGT TCGGTCCTCA GCTCTGGAAA AAAAAAAAAA GACAAAATAA CAAAAAGACC AAAAAAAAAA AAAAAAAAAA AAAACAAGGT AACTAGCACA
 901 CACAACCTAT TGCCGGGCGG TGGTGGTGCA CGCCTTTAAT CCCAGCAGAG GCAGGTGGAT TTCTGTGAGT TCGAGGCCAG GCTGGACCGA CAGCCTCCAA
1001 AACAATACAG AGAAACCCTG TTTCGTAAAA ACAAACAAAC AAACAAACAA AAAACTCAGA ATACGGGCCA GGCAGTGGTG GTGTACACCT TTAATCCCAG
1101 CATTTGGGAG GCAGAGGTAG GCAGATATCT GTGAGTTCAA GGCCAGCCCG GTCTACAAAG TGAGTTCCAG GACAGACTCC AAAGCTACCC AGAGCAACCC
1201 TGTCTCAGAA AACCACTAAA AATCAATCAA TCAATCAACC AATCAATCAG AGTGTGCTGG GTGGTGGTGG TGCATTTCTT TAATCCCAAC ACTTGGAGAT
1301 AGAGGCAGGC AGATCTCTGT GAGTTCGAGA AACCCTGTCC TGAAAAACCA GGAAAAAAAT TCAGACTGAG CCAAGATATG GTTATATTAT ATTCCAGT
```

HUMAN BC200α RNA GENE Length:3903

```
   1 GAATTCTAAG ACAGTGTCAT AGAACATTAG CCCCCCAGTG CAGGACCCTT TTGAACAGAG CTCTTCTGAG CACGAGGCCC TGTCTATGCC TGATGCACAC
 101 CCACGAAACC GACCTTGGCT CCAGCTTTCA ATGCGAACTT CTTACAGTTT AGGTCCTCTG CGAAGTCACT TGCTTCCAGT AGCTCCTATC AGAGCTGAGC
 201 TGAGGCCTTA GACTGGGGTT AAAAGAGAGG TGAAGTTTCC CGCTGCAGCA CTAAACACAT CACTGAAGGC CTACGGAAGT AATCCTTGAG CATCCCCTGA
 301 CTCAGAGCAC TGCCACCGCA TCAGCTTCTT CCAGATGCCA AGAAGCACCT CTGGGTCTTG GGGTTCCGGC TTGTTTGTGT CACAAGTAAT AACTTGTATT
 401 TTATTTGCTT TTAAAAACTG GATATGCTCC CACTTCTCAG CCTCTACAAT GAGACCTGTG TTTCGTGTTT TTCAGTGTGC CTTTGAGAAA GGCCCATAAA
 501 CCTATAAACA CCCTTTTTCT CCCCAAGGTG AGGTTTGACC CTAGTACTTA CTACCACTGT TTACTGGAGA GTCATGTTGA GAACATCTCC TTTAGGCTCT
 601 ATTTCTCATC CTAAGGGATA TATCCAGCCT TCATTATGGA AGGTTTTGAA GGCCTCAAAT AAAATCAAAG TAGAATTTGA ATAAAACATC TTGGCTGGGC
 701 ACGGTGGCTC GCATCTGTAA TCCCAGCACT TTGGGAGGCT GAGGTGGGCG GGTCACAGGT CAGGAGATCA AGAACATCCT GGCTAACACG GTGAACCCCA
 801 TCTCTGCTAA AAATACAAAA ATCAGCTGGG CGTGGTGGCG CGTGCCTGTA GTCCCAGCTA CTCGGGAGGC TGAGGCAGGA GAATCGCTTG AACCAGGGAG
 901 TCGGAGGTTG CAGTGAACCG AGATCACACC ACTGCACTCC AGCCTGGCGA CAGAGAGACT CTGTCTCAAA TAAATAAATA AAACATCTTA AAGAAATAGA
1001 AAATACTAGT AGAGTATTCA CTCTGATTTC AGAATCAATT TGTGGAAATT CAGTTTCACA AGAATGAAAA AAGCCCCCTA AGCAAGGTAG CTTTTTTTTT
1101 TTTTTTTAAA GATGGGGTCT CACTATGTTT TCCAGACTGA TCTCAAACTC CTGGGCTCAA GCAATCCTCC TCCCTCAGCC TCTGGAGTGG CTGGGACTGA
1201 CAGGTACACA TCACCACCAC CCTGCTAGGA TATGGTAGCT TTTTAACCAA AACACACCAC CATGCAATGG AGAGCATGTA ACTTGTTATT TGGTCATACA
1301 AATACATCAG AATAAATTCC ACCTCTCTTG TGGAGCCCTT TTTCCTACAT CTGCACAGAA ACTGGCACTA GCAGGACTGC AACCGAGGCA GACTCGAGAA
1401 TTAGCAGACT CTGGCTCAGG CAGGACAATA GAGCAGCAGG AAATTTCTTT CAAGGGCTTT TGTTGCTGTT GTTGAGATGG AATCTTGCTG TGTCACCCAG
1501 GCTGGAGTGC AGTGGTGCTG TCTCGGCTCA CTGCAATCTC CGCCTCCCAG GTTTAAACGA TTCTCCTGCC TCAGCCTCTC AAGTAGCTGC GATTACAGGC
1601 ATGTGCCACC ATGCCCAGCT AATTTTTGCA TTTTTGGTAG AGATGGGGTT TCACTATGTT GCTCAGGGAG GTCTCGAAAC TCCTGACCTC AGGTGATCCG
1701 CCCACCTCGG CCTCCCAAAG TGTTGAGATT ACAGGCATGA GCCACTGCGC CCGGCTTGAG ACCTTGAGGG GCTTTTTAAC CAAGAGATGG TGTAGGGACA
1801 GAAGCCAGCA AGAAGAACAT GGTGGACTCT CTGTGAGGCC CCAAACCCCA AATCATGAGC CTCCGCCCCC TGAAAGTATT CCAGGATCTC TGAGGGAGGT
```

1901 TTTGGGAGGG GGCCAAGGTC CTGCAATTGG GTGGGTGGGG CTCCAAGCCA TTGTTCTTTG CACATTCATG CCAATGTGAC CTAATTGGCT AAGCACAGGC

2001 TGGAGAAGCT AGATAAACCT ATTTAAATTC AGAGCCAAAA GACACTCAAG GATGTTTTCT GAGGGGGTGT GAGGGTTGGA AAATCATCCC TATATTTGAT

2101 TTATCTCCCA CCTCATGTCC TGAGTCAATT CCTTAGTCTT AGATGAGGGA GACGACTTGG GAGTCATCCT TGTTTTTGAT GAGCTATATA ACCCTATGGC

2201 CAGCAGAGGG AAGTACTGCA TTTCAGAGCG ACAATTTGAG ATCTATGAAA GAATTTCAAT CGAGAATAAG AGGCCGGGCG CGGTGCTCAC GCCTGTAATC

2301 CCAGCTCTCA GGGAGGCTAA GAGGCGGGAG GATAGCTTGA GCCCAGGAGT TCGAGACCTG CCTGGGCAAT ATAGCGAGAC CCCGTTCTCC AGAAAAAGGA

2401 AAAAAAAAAA CAAAAGACAA AAAAAAAATA AGCGTAACTT CCCTCAAAGC AACAACCCCC CCCCCCCCTT TTCATATTCT TGCCTGGAAG AAAGGCCTTG

2501 CTTTCCTCAG CTTCTTAAAG CTGGGAGAAG TTAAAGCCAT TCTGAAATGC TGCTGCTACT GCTATATGTG GGGGAGAAAG CAAATTTTCT TTTTACTTTT

2601 TTAATCAGTA TTTATTTTAA GTTCGAGGGT ACATGTGCAG GATGTGCAGG TTTGTTACAT AGGTAAACGT GTGCCATGGT GGTTTGCTGC ACAGGTCAAC

2701 CCATCACCTA GGTATTAAAC CCAGCATCCA TTAGCTATTC TTCCTGATGC TGTCCCTCCT CCCACCCCAC CAACAGGACC CCAAATTTTC TTTCTTTTTT

2801 TTTTTTTTTT TTTTTTGAGA TGGAGTCTAG CTCTGTCACC CAGTCTGGAA TGTAATAGCT CAATCTCGGC TCACTGTAAC CTCCATCTCC CAGGTTCAAG

2901 TGATTCTCCT GCCTCAGCCT TCCAAGTAAC TGGGATTACA GGCATGCACT ACCATGCCTG GCTAATTTTT GTATTTTTAG TAGAGACCGG GGTTTCACCA

3001 TGTTGGCCAG GCTGGTCTCA AACTCCTGAC CTCAAATGAT CCGCCTGCCT CAGCCTCCCA AAGTACTGGG ATTACAGCTG TGAGCACCGT GCCCAGCCTG

3101 CCAAATTTTC TTATTTCCAT GGCCTTGAGC AAAATTATAC CATCCTAGAC AAACACACAC ACACACACAC ACACACACAC CATTCTCCCC CTCTTGCTGT

3201 CTTTCTCAGA TATGAGTCTC TCCTCCAATG GCCATTTTAA GCTTGCTCTT TTGCTCTGTA ATACAAGGCC TGTTCCCTGT TAAAAGGAAT GTCAGGCATC

3301 TGTGAGCAGG TAGGGAAAAC ATTGCAGGAG GCAAGTAAAC TTCTCAGCAT ACCTGCCATC ATTCATACCT TGAAGGCAGG TCAAGCCTGT AGCGTCTGAG

3401 TCTTCCATCC CTGAAGCACC AGCAGCTCTG ACTAGAGGTC CTGTCTCCCA GCTGTGGCCC CAGCTCTCCC TGATGGTGAC AGTTGCGACC CAAACTTGCT

3501 CCTCCCCTAT ATCTGTAACT GAGCCCAGCG TACATTTGCC TCATGGGTGC CATTGGACTT CCTTTGTGAG CCATGTTTTC TTTCACCTTG TCCCAATCCA

3601 CAGTTAACCA AATCTGGCCC AAATTATGTG CTGGCTGGTG TTGGGCAGCT GTGAGACCCC TCCTTGCCCC ATCTCTGTCT TCTAGGCTCT TCTGAGAACC

3701 AGTGTCCTTT GAGATCTCAG ATGGGATTTA GAGGAGGGTT TCCACCCAGA CCACTTTCCA CAAATTTCAC TTCTGCACTT TCATCTGGTA GGGTGTGGTT

3801 TTCAACTAGA AGCCCAAATC GAGGAAGAAA AAAATTCATT CACCATTCAG GACAGGTGGT TCACATTTGA ACCAGGCAGA AATAGCTGAA AGAAACTGAA

H

3901 TTC

HUMAN BC200β RNA GENE Length:5975

```
   1  TCGATCCCAA GAAAGGGACT GAGTTGCCTC ATAGGAAGGA AGTCACTTAC GGAATATGTC AAGGCCCACA GACAGTAATT AGGCTAATTA TTAGAAGGAG
 101  AAATATTCTT AGTATAGTAG GTTCATTTGT TCATTCATTC AACAAATATT GAGGTCATAC TATTTTCAAG ATACTATACT AGCCACTGAG GGATATATGG
 201  GGAGCAAAAG CAGACAAGGC CCCTGTCCTT AAAGGAGCTC TCAAGTATAA TTGGAGGAAA TATGTAAGAT AATAAAGGAT GATGTTATTA TTTAATCAAA
 301  TTTATATAGG GCCTATGTTG TGCCACGTAC TGCTCTAAGC ATTTTACATG TAGTTGACCC ACACAACAAC CATATGATGT AGGTACTGAT AATATCTCCA
 401  ATCTACAGAT GACGAGCCTG ACACACAGAT ATGTTAAGTG ACTTGCTCAA GGTCACCCTA CTTGTAAGTG GCAGAGTCCA CATGCTTACC TCCCATGCCA
 501  TACTGCTTTG GGATGATAAT GGTACAGGTA CAGAGTGACA GATCATGACT CTCACAACTT TTTCTCAGGT GCATTTAGTG CCACTAGTTG TTCAGAACTT
 601  ATTAACTCAA TAAACATTTC CTTAGGGTCC CACTGTGTGG CTGGCACTGC ACTAGGCAAT AAGGATCTAC CGATGAACAA GACTGATTTA TTCACTGCCC
 701  TCAATGGAGC CTGCAGGCCA GCAGAGAAGG AAGATATTGA ACAAGTGACA TGAAAAGGGA AGCACAGAGA GTAATGGAAG CAGATAACCT GGGACTCTGG
 801  CCCAATTTCA GGAGGCATAT GTCTCCATCA CTCAGGTTGT CCTTTGCCTC AGGCCCACGG GGAAGCAGAC CCCATGGGAC CTCCTCCTGT TAGACAAACA
 901  TGGCACTATA GGTTATCTGT GTGAGATGGA GGGAACACAA GGGAAATTGT CTAGGGATCA CAAAAGAGAT GGAGACAGGG ATTCCCCCCA TGACACCTAC
1001  ATCTGAGAAG ATAGCATCAA TTGCATGGTG GTATCCATGG TGAAAGGTGG TATTAGTGTA ATCAGGCAGC CATGGAACAC CTTCACAGCA GCTTTGGGGC
1101  CACAGATATC TCACTGTGGG AGGCAGAGAA GAAACAGAAG AGACCTTCAG ATGCCATATG GAGAGAGCTG TGTCCATGAG CTCTTACGAG AACGTGCAAA
1201  AGAGATGACT TTGGTGGGGG ACTAAGGGTC CCACAGTAGC AGGTAAGCAC AAAATCAGCA AAACATCATA CTCCTATTAT TCTGATGCAA TTTTTAAACC
1301  AGCTAGTGTG ATCTGGGAAA TGGAAATTAG AACAACAAGG GTAACTAAAA AGTTCTGCCT TCATGCTCCA AAAACCAATT ATACAATTAG AGGATGGGAA
1401  GCTGTAGTGG GTTGGTGGTC CTCAAAAATA TACGTCCATG CCCCAATCCC TGGGGCCTGT GAATGTGATC TTATTTGGCA AAAGGCTCTT TACGGGTGTA
1501  ATTAAATTTA GGATCTCAAG ATGAGGTTAT CCTGGAGGGT CCTAAATTAA ATGGCAAGAA TCCTTGTAAG ATACACACAA AAGAGAGACA AAGGGAGGAG
1601  AGGAGAAGGC CAGAGATAGG AATTATACAG CCATAAGCCA AGGACTTAGA ACCACCAGAA ACTGGCAGAA GCAAAGAAGG ACTCTTATAG AACCTTCAGT
1701  GAGTGTGTGG CCCTGCTGAC ACCTTCATTT TAGAGCTCTT GGTTCCAGAA CTGTGAAAGA GAATACATTT CTGTTATTTT AAGCCACAAA GTCTGTAGTA
1801  ATTTGTTACA TCTGCCCAGG GACAGTAATA TAGGGACCCA TGGCTTGGCA ACGATTTATG TAAAAAATAT CTGGGCAAAC AGCATGAAGT GACTGATAAA
```

1901 GAAAGCAAAT GCAGGGCCAG GCCTAGTGGC TCACACCTGT AATCCCAACA CTTTGGGAGA CCAAGGCGGG AGGATCACTT GAGGCCAGGA GTTCCAGACC

2001 AGCCTGGCAA ACATGGCAAA ACCCAGTCTC TACTAAAAAT ATAAAAATTA GCTGGGAGTG GTGGTGCACA CCTGTAATCC CAGCTACTTG GGGGGCTGAA

2101 GCAGGAGAAC TGCTTGGGCC CCAGAAGTGG AGGTTGCAGT GAGCCAAGAC TGCACCACTG GACCACTCCA GCCTGGGTGA CAGAGCGAAA CCCTGTCTCA

2201 AAACAACAAC AACAAAAATA AAGCAAATGC AGTCTTTGGC TGCCACAAGA GAAACAGAGT TAATAGACTG AGGAAGCTAA TCATTCTGCC ACGTAGCTGC

2301 AATATTGCAC TCTGCTCTAA GTGCCACCTT TTAAGAGAGA CATTGATGAA TGGGAGCCTA GCCAGAGATA GGTGTCTGGA ACTGTGTTGG TTCTGGAGAG

2401 CTTGCCATAC AAAGGATGAG CTGGTAGCAC TGGAGATGGG GATCTTGGAA TAGAAATCAA TGAACATGAA GGTAACCTTG GAATATGTGA AAAGCTATCT

2501 TGTCAAAGAG GAAGTAATCT TGTTTTGTGT TGCTCCAGGA GGCAAAATTG GGTTTAGCAA GTGGAAAGTA AAACAGAAGA AAAAACTAAT GACTGTTGAT

2601 TACAGAAATT CAAAACATTG GAAGTATTTG AAATGGGGAA AAGTGCTGTT TCCTAATTCC AGAGACAGGA CATCTGGTCC TGCCTCTGTA GTTAATTGGT

2701 TGTGTGACTC TGGACATATC ATCTCCCCTT CTTGGGACCT CAGTTTTCTT CTATGTAAAA TTATGGGGTT AGGTTGGATG CTTTGGGAAG GCATTGTTTT

2801 TTTCCTTTTG GTTGTTTCCT ATTTTATGGT TTAATAAAAC TGTTAAGATG TTTTACAATT TCTGAGTGCC TAGTATATGC TAAAGCTAAA ATCTGGGCTA

2901 GGTGTTTTGT TTTGTTTTTT GAGATGGAAT TTCGCTCTTG TCGCCCAGGC TGGAGTGCAG TGGTGCAATC TCAGCTCACT GCAATCTCTG CCTCCCGGGT

3001 TCAAGCGATT CTCCTGCCTC AGCCTCCCGA GTAGATGGGA TTACAGGCGC CCGCCACCAC GTCCGGCTAA TTTTTGTATA GTTAGTAGAG ACAGCATTTC

3101 GCCATGTTGG CCAGGCTGGT TTCAAACTCC TGACCTCAGG TGATCTGCCT ACCTCGGCCT CCCAAAGTGC TGGGATTACA GGTGTGAGCT GCCATGCCTG

3201 GCCTGTGCTA GGTGTGTTAA CATGCATGAT CTCATATAAC TCTCTCCACA CTACAGGTTA GGTATGTTCC CCCATTATGA ATTGTTCTTG GTGTCAACTT

3301 GACTGGATTA AGGGATACTC AGGTAACTAG CAAAGCATTA TTTCTGGGTA TATCTGTGAG GGTGCTTACA GGAGAGATTG GCATTAGAAC CAGTGTACCT

3401 AAGTAAGGAA GATCTACCCT TACCCATTGT GCGTGGGCAT CAGCCAATCT GCTCAGGGTG CAGAACAAAA AGGCAGAGGA AAGGAGAATT CACTCTCTTT

3501 CTTCTGGAGC TGGAACACCC TTCATCTGCT GCCCTTGGAC ATCAGAACTC CAGGTTTTCT GGTCTTTGGA CTCTAAGACT AGCACCAATG GCTCACCAGG

3601 TTTTTAGGCC TATGGCTTCA GACTGAGAGT TATATCATTG GCTTCCCTGG TTCTGAGGCC TTTGGATTTG GACTGCACCA TTCTACTTGG CTTTCCTCAT

3701 TCTCCGGCTT GTAGATGGCC TGTCATGGGA CTTCTCAGCC TCCATAATCA CCTGAGCCAA TTTCCCTGTG AATCTCCTCT CCATGCCTCT CTCTCTCTCT

3801 CTCTCTCTCT GTCAGTTCTG TCTCTCTGGA GAAACCTAAT ACATACCCAT TTTACAGATT AGAGCACTGA GATTGAGAGA GATTGTGACT TGCCCAAAGT

K

3901 CATATAACTT GTTCATGGGT GAGCTAAAAA TTGAACTCAG ATCCCCCTGC TTTTAGTATT CTTTCCACCA TATCATACTG TCTCCCCTAC ACAATAGTGA

4001 CCCTTTGCCA CACCTGAATT TAGCAAATTA TACAATTCAT TGTCAAAGTA CTCCTCAGAA ATCATTTACC AGGCAATATT AACTATCTCA AACACCCCAA

4101 AGAATCAGAA CATTTCTGAG CAGACAAAAT GGTTGGTGCC AAAGCCTATG CGTTAAAGCC CATAAACTTC ATTCAGAAAA ATTCTTAGGG CTGCCTCACC

4201 CAGAGCCTGT CCCTCATCAT TTCGTATACA GTGTCTGTTC TTCCTTCCCG AAGAGAAAGG CAAGCTGGAG GGCTGGTCTG CAGCATCCTG GGAGCCTGAG

4301 ACAGCCTGGG TTGTTAGGCT GTGGAGCATG GAGGCAAAGG CAAAAAGATT AGCAGTGTCC CGTGGAGTCT CCACAAAAGG GGAAGATAAA CAGCAAGGTC

4401 AATGAGTCTA GCCCTTTGGA ATTCCATGCA TGAGAGGCTC AACACAGGAA GCAGGTCAGA GAGGTCTAGA GGAATATCTT GAACCAGTGA CTACGCAGCC

4501 AGCTGCAGAG GTTTTTACAT GGTTCCGGGG ACTCTGGGGT GGGAAGAAGA GAGCAGTACC CTGAGAGTGC AGTACAGTCC AAAAGTATAC ACTAGGTTCT

4601 GATATACTAC TCTGGATTTT TAGTCAGTTA CATGCTATAA TTTAAAAAAA ATCTGTGGGC CGGGTGCAGT GGCTCACGCC TGTAATCCCA GCTCTCAGGG

4701 AGGCTAAGAG GCGGGAGGAT AGCTTGAGCC CAGGAGTTCG AGACCTGCCT GGGCAATATA GCGAGACCCC GTTCTCCAGA AAAAGGAAAA AAAAAAACAA

4801 AAGACAAAAA AAAAATAAGC GTAACTTCCC TCAAAGCAAC AACCCCCCCC CCCCCGCTTT AAAAAAAAAA AAATACAAAT AAATCAGCCG GGCGTGGTGA

4901 CAGGCGCCTA TAGTCCCAGC TACTCGGGAG GCTGAGGCAG GAGAATGGTG TGAACCCGGG AGGTGGAGCT TGCAGTGAGC TGAGATGGCG CCACTGCACT

5001 CCAGCCTGGG CTACAGAGCG AGACTCTGGC TCAAAAAGAA AAAAAAAAAT CTGTGGGGGA AGTCAAGAGG AAGAAGGTAG ACATGGACCA ATAGGCGTAA

5101 CAAAAATTAA ATATTGACAA TGAGAGCAAG GGAGCTGACA GGAAGACTAT AGAAAGAAAT ACGATATCCT CCTAAAGCTA AGCAGTGTGA AGACCCAGAA

5201 AGGCTGGCAC TACTGTACTT ATTGCCGAGT CCTTTAGTGG ATATCCATAA TGATGACTCA GCGTTATTGA GAAGGTGAGG TGGTATTCTG TGCCTTCTGC

5301 ATATGAAGGC AGAGATGAGT GCAACATATT ATAAAAAGAC TTCCACCCAC GATGACCTGA AGAAAAAAAT TTAATGCCCC AAAGAATAAT TGCAGTTGTC

5401 TAGCAAATGC AACGATGTAG AAATTGACAG AGCAGAATGA AAGAGATAAT GCTTTAAAAA TAATATTAAA GGCAGCTGTT AAGGAATGTA GTAAATGTGG

5501 GAGAAGGCAC TCAATAAATA CGTTTAAAAA ATTAAACTCA GTTTGGACCC TGCCTAGCTC CCATTCTGGA GAGGTATCTA AATAAATGAA TGGGTGACTG

5601 AATGTATGAA TGAAAATTAT CTGGCCTGAG GTCCCCCTTT CACTAGGGTT GTTCTATGGT TCTTGGAGCT AATGCAGTGT CAAGGTGTTG GAGATGGAGG

5701 ACACCTGTCT TATCTTCAAT CATCATCCAC ACACATTACC ACTCAGGAGG TCCATTATCC CTATCTGCAT CATGTGTTCC AGGGCCAGAA GCTGTTCTGT

5801 GTTCTTGCCA TGTTGAGGAA TGGTTTCTGC CAATGCTGGC TAGGATGCCA TTTGCCTTGA AGCACCCCTC TATTGTCCTC TCTGTTCTTG TTTCCCTGAA

5901 GGTGCTGCTC AGCAGAACAG TGGTCTGACC TGATCTTTGC ACTAAAAGAT CCTTTCTCAT TTCTCCTCCC TGCAG

M

HUMAN BC200γ RNA GENE Length:538

```
  1  GATCTCATTG TTCAATTCCC ACCTATGAGT GAGAATATGC GGTGTTTGGT TTTTTGTTCT TGCGATAGTT TACTGAGAAT GATGGTTTCC AATTTCATCC
101  ATGTCCCTAC AAAGGACATG AACTCATCAT TTTTTATGGC TGCATAGTAT TCCATGGTGT ATATGTGCCA CATTTTCTTA ATCCAGTCTA TCATTGTTGG
201  ACATTTGCAG TGGCTCACGC CTGTAATCCC AGCTCTCAGG GAGGCTAAGA GGCGGGAGGA TAGCTTGAGC CCAGGAGTTC GAGACCTGCC TGGGCAATAT
301  AGCGAGACCC CGTTCTCCAG AAAAAGGAAA AAAAAAACAA AACAAAAGAC AAAAAAAAAA AAATAAGCGT AACTTCCCTC AAGGCAAACC AATCCAATCT
401  TTTTAACACA TTTTTATGTA TGTATGTATG TATGTATTTA TTTTTAGACA TAGGTGGGCC ACCGAAGTTC TTTGACTTGC TATTTATTTC TTTCCACCAT
501  TATGTCTCCA CCTCTCTAGT AGCGCTGAGC CAGTCGAC
```

CHIMPANZEE BC200β RNA GENE LOCUS Length:595

```
   1  TTTTTACATG GTTCCGGGGA CTCTGGGGTG GGAAGAAGAG AGCAGTACCC TGAGAGTGCA GTACAGTCCA AAAGTATACA CTAGGTTCTG ATATACTACT
 101  CTGGATTTTT AGTCAGTTAC ATGCTATAAT TTTAAAAAAA TCTGTGGGCC GGGTGCAGTG GCTCACGCCT GTAATCCCAG CACTTTGGAA GGCCGAGGCG
 201  GGCGGATCAC AAGGTCAGGA GATCGAGACC ATCCTGACTA ACACGGTGAA ACCCCGTTTC TACTAAAAAA TACAAATAAA TTAGCCGGGC GTGGTGACAG
 301  GCGCCTGTAG TCCCAGCTAC TCGGGAGGCT GAGGCAGGAG AATGGTGTGA ACCCGGGAGG TGGAGCTTGC AGTGAGCTGA GATGGCACCA CTGCACTCCA
 401  GCCTGGGCTA CAGAGCGAGA CTCTGACTCA AAAAGAAAAA AAAAAAAAAA TCTGTGGGGG AAGTCAAGAG GAAGAAGGTA GACATGGACC AATAGGCGTA
 501  ACAAAAATTA AATATTGACA ATGAGAGCAA GGGAGCTGAC AGGAAGACTA TAGAAAGAAA TACAATATCC TCCTAAAGCT AAGCAGTGTG AAGAC
```