*Nucleic Acids Research, 1993, Vol. 21, No. 6 1351–1359*

# Proposed roles for DNA methylation in *Alu* transcriptional repression and mutational inactivation

Wen-Man Liu[1] and Carl W.Schmid[1,2]*
Departments of [1]Genetics and [2]Chemistry, University of California – Davis, Davis, CA 95616, USA

Received January 11, 1993; Revised and Accepted February 22, 1993

## ABSTRACT

**Methylation at CpG dinucleotides to produce 5 methyl cytosine (5me-C) has been proposed to regulate the transcriptional expression of human Alu repeats. Similarly, methylation has been proposed to indirectly favor the transpositional activity of young Alu repeats by transcriptionally inactivating older Alu's through the very rapid transition of 5me-C to T. Both hypotheses are examined here by RNA polymerase III (Pol III) *in vitro* transcription of Alu templates using HeLa cell extracts. A limiting factor represses the template activity of methylated Alu repeats. Competition by methylated prokaryotic vector DNA's relieves repression, showing that the factor is not sequence specific. This competitor has no effect on the activity of unmethylated templates showing that the repressor is highly specific toward methylated DNA. While methylation of a single pair of CpG dinucleotides in the A box of the Poll III promoter is sufficient to cause repression, methylation elsewhere within the template also causes repression. The repressor causing these effects on the Pol III directed transcription of Alu repeats is thought to be a previously reported, repressor for Pol II directed templates. Young Alu repeats are transcriptionally more active templates than a representative older Alu subfamily member. Also, younger Alu's form stable transcriptional complexes faster, potentially giving them an additional advantage. The mutation of three CpG's to CpA's within and near the A box drastically decreases both the template activity and rate of stable complex formation by a young Alu member. The sensitivity of Alu template activity to CpG transitions within the A box partially explains the selective transpositional advantage enjoyed by young Alu members.**

## INTRODUCTION

### Possible regulation of Alu transcription and transposition

Nearly one million Alu repeats in human DNA result from the retrotransposition of an RNA intermediate (1,2). Alu repeats are, with few exceptions, readily transcribed *in vitro* by virtue of their internal Pol III promoter (1–9). Surprisingly, the corresponding *in vivo* transcripts are sparse, considering that there are one million potentially active transcription units (10–12). The relative paucity of these transcripts might result from Alu's retrotransposing into inactive chromatin domains or into regions lacking positive regulatory sequences or from a combination of these and many other conditions. As discussed below, DNA methylation may also repress Alu transcription.

Several independent investigations show that Alu repeats belong to distinct sequence subfamilies which appeared at different evolutionary times in the ancestral human lineage (for review, see 13–14). Here the youngest subfamily is called PV, the next oldest is Precise, and the next oldest, which includes the majority of members is called the Major Subfamily. The extent of divergence within each subfamily's membership correlates with the time of its evolutionary appearance. Also, CpG dinucleotides show a very rapid transition to TpG (or equivalently CpA) which is indicative of 5me-C abetted transitions. Recently inserted Alu's, regardless of their subfamily affiliation, have almost all of their CpG's unmutated (14). For simplicity, we refer to PV and Precise Subfamily members investigated here which have most CpG's intact as being young. Major Subfamily Alu's that do not have intact CpG's are referred to as being old.

The question of whether recent Alu inserts are encoded by one or many sources is being actively debated (13–16). One serious difficulty with recruiting sources from the dispersed membership of the Alu family is the additional requirement to silence older Alu's so that recently inserted young Alu's are derived from source sequences that have mostly intact CpG's. In the single master gene model this difficulty does not arise as the source might be hypomethylated which would maintain its CpG residues (13). As a solution to this difficulty in the multiple source gene model, we imagine the possibility that rapid 5 me-C to T transitions quickly place older Alu's at a selective disadvantage in the retrotranspositional pathway. These mutations might affect any of the many steps in this pathway (14) including transcription, the possibility investigated here.

### Regulated methylation of Alu repeats

The young PV and Precise Alu Subfamily consensus sequences are very rich in CpG dinucleotides (9%) as compared to total human DNA (1%) and recently inserted members of these two subfamilies retain all of their CpG residues (13,14). In somatic

* To whom correspondence should be addressed

tissues these CpG's are highly methylated; for example, in spleen, Alu CpG methylation approaches 100% (17). In contrast, Alu CpG's are completely unmethylated in sperm and other selected tissues associated with the male germ line (18). These developmental differences in the methylation pattern of Alu elements may also regulate their transcription.

### Methylation affects Pol III transcription *in vitro*

Two distinct classes of Pol III transcribed genes have intragenic promoters: 5S RNA genes belong to type 1. Alu repeats, tRNA genes, and adenovirus VA1 RNA genes and others belong to type 2 (19). Methylation does not affect the transcription of type 1 genes (20). The transcription of tRNA genes assayed in Xenopus oocytes is inhibited by methylation (20). Previously, effects of methylation on the transcription of Alu repeats and tRNA genes were not detected in *in vitro* transcription assays (5,20). However, in a recent study (21), the effects of methylation on the *in vitro* transcription of the VAI RNA gene were reliably detected at lower template concentrations than those employed in these previous studies. Methylation inhibits transcription of this gene both *in vivo* and *in vitro*. These findings imply that either a methylation sensitive repressor is present in limiting amounts or that positive transcription factors have a lower affinity toward methylated templates. In the case of the VAI RNA gene, the inhibition of transcription has been attributed to methylation interfering with complex formation by positive transcription factors (21). A repressor that binds specific to methylated DNA has been identified for Pol II transcription units but its effect on Pol III templates has not been reported (22, 23). For these reasons, the effect of methylation on Alu transcription *in vitro* at low template concentration is reinvestigated here. This same investigation provides an opportunity to compare the transcriptional activity of old and young Alu templates.

## MATERIALS AND METHODS

### DNA preparation and methylation

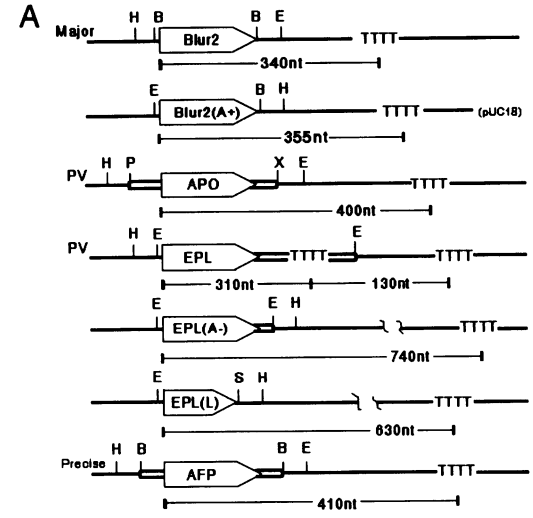
Closed circular DNA was prepared by the alkaline lysis method and purified by CsCl-ethidium bromide centrifugation (24). DNA concentration was determined by uv absorption spectroscopy and in every case, the relative concentrations of different DNA templates were verified by agarose gel electrophoresis stained with ethidium bromide.

Methylation was performed with CpG methylase (M. SssI, NEB) or FnuDII methylase (NEB) using at least a two fold excess according to the manufacturer's protocols. Complete methylation was shown by resistance of the plasmid to digestion with methyl-sensitive restriction enzymes HpaII or BstUI. In all cases, the unmethylated control template was subjected to mock methylation excluding only the methylase. Templates were repurified by phenol extraction and ethanol precipitation.

### Cell culture and extract preparation

HeLa cells (ATCC CCL 2, 1990) were grown in spinner flasks at 37°C in SMEM (Minimum Essential Medium, GIBCO) containing 5% calf serum; F9 cells (ATCC CRL 1720, 1990) were grown in DMEM (Dulbecco's Modified Eagle Medium, GIBCO) supplemented with 10% fetal bovine serum, 5% calf serum and 0.35% glucose. They were grown to 4 to $6\times10^5$ cells per ml prior to harvesting for extract preparation (4).

Nuclear and S100 extracts were prepared from HeLa and F9



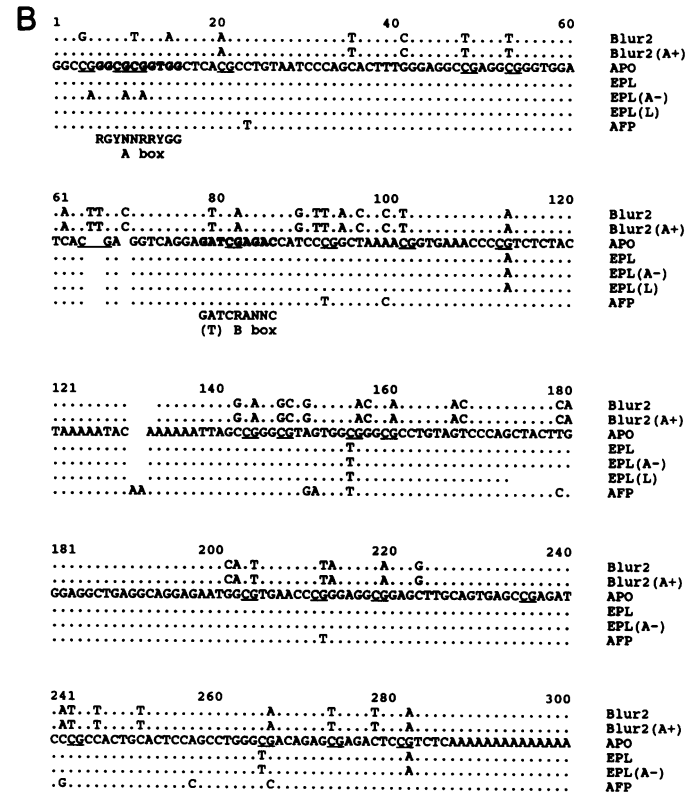




**Figure 1.** (A) Restriction maps of Alu templates show their orientation with respect to the pUC19 vector multicloning sequence (H is HindIII and E is EcoRI) and the first terminator (indicated by TTTT) supplied by pUC sequences. The single line indicates flanking vector sequence; the double line indicates eukaryotic DNA. The subfamily identities of several clones are written on the left. EPL includes an endogenous terminator in its flanking sequence and Blur2(A+) was force cloned into pUC18 to use the same terminator as its parent clone Blur2. Other restriction sites are BamHI (B), PstI (P), XbaI (X), and SmaI (S). The predicted size of the major product is indicated in nucleotides for each template. (B) Base sequences of Alu templates are compared to the APO Alu template with a dot indicating an identity and a letter indicating a base substitution. CpG dinucleotides are underlined and the A and B boxes are presented in bold type. Consensus A and B box sequences are written blow the Alu sequences.

cells according to the method of Dignam et. al. (25). Protein concentrations were determined by the coomasie method (26) using bovine serum albumin as the standard.

**Alu templates investigated and assignment of their transcripts**

A total of ten Alu templates were investigated during the course of these studies, of which seven are depicted in Figure 1A and B. The following discussion follows the order presented in Figure 1A. **Blur2** is cloned by BamHI linkers adapted to a renatured Alu devoid of other flanking sequences (27). The Blur2 sequence resembles the Major Alu Subfamily. PCR mediated synthesis, as described below, was used to construct **Blur2(A+)** which has a perfect PV Subfamily consensus A box. **APO Alu** is a 500 base pair XbaI−PstI restriction fragment (28,29). This particular Alu is young by the criteria that it is not fixed in the human population and differs from the PV consensus sequence by only a single base change. EPL Alu is thought to be the ancestral source of the PV Alu Subfamily (30). This sequence was originally subcloned as a restriction fragment having about 400 base pairs of 5′ flanking sequences. The 5′ flanking sequences were removed using PCR strategies outlined below and oligo 928 which contains an EcoRI site as a primer. The resultant PCR product is cloned as an EcoRI fragment to give **EPL** (Figure 1A). Methylation is shown to have an identical effect on the transcription of the parent clone with flanking sequences and the PCR product daughter clone, called EPL. Consequently, no additional studies were performed on the parent clone in this investigation. PCR mediated synthesis, as described below was used to mutate three CpG residues in the A box of EPL to give **EPL(A−)**. **EPL(L)** is generated by cutting the EPL Alu at its internal AluI restriction site and cloning the product band into EcoRI−SmaI cleaved pUC19. **AFP Alu** is a member of the Precise Alu Subfamily and is cloned as an 890 base pair BstYI restriction fragment into BamHI cleaved pUC19 (31,32). This particular Alu is young by the criterion that it is fixed in the human population but absent at the orthologous sites in gorilla and chimpanzee DNA's. The Major Alu Subfamily member located immediately 3′ to the human α-1 hemoglobin gene, **3′α-1 Alu**, was subcloned as a 400 base pair HincII−EcoRI restriction fragment into pUC18 (5).

Two constructions shown in Figure 1A, Blur2(A+) and EPL(A−), were generated by a PCR approach using the following primers. Oligo 810: 5′-AAAGCTTGGCCGGGCG-CGGTGGCT-3′, 811: 5′-AGAATTCTTTTTGAGACGGAGT-CTCGCT-3′, 928: 5′-TTTGGAATTCGGCCGGGCG-3′, 1440: 5′-GGCCAGGCACAGTGGCTCAC-3′ were custom synthesized and a vector forward (−40) primer was purchased from USB corp. PCR conditions are 0.025 mM for each dNTP, 1×Taq buffer (Invitrogen), 2 units of Taq polymerase, 300 ng of each primer and 50 ng Alu template, in a final volume of 100 μl. This mixture was incubated for 0.5 min at 95°C, 0.5 min at 45°C and 1.5 min at 72°C for 35 cycles, followed by an additional 10 min at 72°C.

The EPL(A−) PCR product which was generated by using oligo 1044 and oligo 811 on the EPL template, was cloned into TA vector (Invitrogen Inc.). We were concerned that EPL(A−) used a different pUC terminator than most other Alu constructs investigated here (Figure 1A). As a control on the possible effects of different terminators, we also constructed and investigated the opposite orientation of EPL(A−). The EcoRI fragment from the TA construct was subcloned into pUC19 (Figure 1A). The opposite orientation of EPL(A−) was derived by force cloning this product into pUC18. The pUC18 clone uses the same proximal pUC terminators of most of the other templates investigated here (Figure 1A). Because the level of transcription was not different between these two template orientations, the pUC19 clone shown in Figure 1A was selected for all further studies reported here and the pUC18 clone is not further investigated.

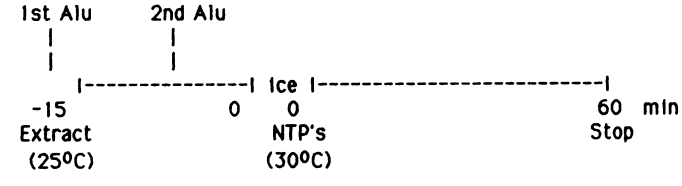
Two PCR reactions were used to generate Blur2(A+) (Figure 1A). The first PCR reaction included oligo 810, the forward vector (−40) primer , and Blur2 template. The second PCR reaction used oligo 928 and the −40 primer with 5 μl of the first PCR product. The second PCR product was digested with EcoRI and BamHI, and force cloned into pUC18 to have the same terminator as its parent clone, Blur2.

Sequencing was performed to determine the orientation and authenticity of the Alu templates (Figure 1A). Alu specific transcripts were identified by comparing the transcription products of the various clones to a pUC control template. In each case, a major product corresponding to the predicted transcript length resulting from termination in the first run of four T's was observed (Figure 1A). As discussed in Results, minor bands resulting from a readthrough of the first terminator are observed in each case. The lengths of these minor bands are also consistent with termination occurring in runs of four or more T's present within the flanking pUC sequence. These results suggest that the transcripts are Pol III products (19,33). Previous studies verified the transcription of Blur 2 by Pol III (3). Also, we find that the major transcripts of the Blur 2, APO, EPL and AFP Alus are resistant to low concentrations of α-amanitin; again, verifying that these transcripts are Poll III directed (7).

***In vitro* transcription assay and quantitative analysis**

Transcription assay reaction mixtures (50 μl) typically contained 10 to 20 μl of nuclear extract (about 5 to 10 μg protein), DNA template, 8 mM HEPES (pH 7.9), 0.11 M KCl, 4 mM $MgCl_2$, 0.7 mM 'DTT, 0.8% glycerol, 0.5 μM α-[$^{32}$P] CTP (Amersham), 25 μM unlabeled CTP and 600 μM each of the ATP, GTP and UTP (Pharmacia). The total DNA concentration was adjusted to a fixed concentration (from one to four μg per assay) in each experiment by the addition of pUC19 DNA. The reaction mixture was incubated at 30°C for 60 min and transcription was stopped by the addition of 50 μl of 0.4 M NaCl and 0.2% SDS. RNA was purified by successive extraction with phenol- chloroform and chloroform alone, and precipitated with 250 μl of 100% ethanol and 5 μg of carrier RNA (yeast tRNA, Sigma) (4).

Order of addition experiments were essentially performed as described above except for the inclusion of a preincubation step as depicted below (19,34,35). The 1st Alu template was incubated with extract in the absence of NTP's at room temperature, and the 2nd template was added at various times during the following 15 minute interval. To synchronize reaction times, assays were stored on ice for a maximum of five minutes prior to adding the NTP's (34).



To analyze the *in vitro* transcripts, RNA pellets were dissolved in 10 μl deionized formamide, denatured at 90°C for 5 min, and electrophoresed in 5% polyacrylamide 8 M urea gels. DNA sequence ladders of M13 mp18 were used as size markers for products having lengths of less than 600 nt. The identity of longer products was assigned using the maps in Figure 1A. The gels were autoradiographed at −70°C using an intensifying screen (Du Pont). A Fuji BAS 1000 Bio-Imaging Analyzer was used to measure radioisotope incorporation into RNA transcripts.

## RESULTS

### Effect of template concentration on methylation induced repression

In preliminary experiments, methylation inhibited Alu transcription at low but not at high template concentrations. Crude extracts contain limiting amounts of both positive and negative factors so that the effects of simultaneously changing both template and total DNA concentration are quite complex (35). In subsequent experiments, total DNA is kept constant by the addition of vector DNA (Figure 2).

At low template concentrations (10 to 50 ng total DNA), there is a 2.7 fold inhibition of the major 400 nt. transcription product from the APO Alu due to methylation, but at high concentration (i.e. one μg), inhibition is undetectable (Figure 2, Table I). Nuclear extracts are used in these experiments while S100 extracts were previously used to study the effect of methylation on Alu transcription (5). As shown below, a methylation sensitive repressor is present in limiting concentrations in nuclear extracts and, as might be expected, this repressor is more difficult to detect in S100 extracts (data not shown). For clarity, we shall henceforth refer to this transcriptional inhibition as being due to repression (see below).

{An interesting detail in the data of Figure 2 deserves additional attention. While the intensity of the major transcript is repressed by CpG methylation of the APO Alu template, the intensities of the 555 nt and 750 nt minor product are enhanced by methylation (Figure 2). These minor products result from a read through of the first terminator, consisting of four T's located in the pUC flanking sequence, followed by termination at a downstream sites in pUC which also consist of four T's (Figure 1). The major upstream terminator is immediately preceded by two CpG dinucleotides and within ten nucleotides is followed by an additional two CpG's. Evidently, methylation of these CpG's interferes with termination at the major site; thereby enhancing transcriptional readthrough for subsequent termination at the minor sites down stream. As evidence for this interpretation, methylation of the APO Alu template with FnuDII methylase (see below), which does not target the first pUC terminator, does not enhance the intensity of the minor product. Also this effect is a peculiarity of the first pUC terminator which resides near several CpG's and is not observed for additional Alu templates described below which use terminators supplied by the endogenous human DNA flanking sequence. As demonstrated below, a limiting factor recognizes methylated DNA and possibly this same factor causes this subsidiary effect on termination. Since there is no evidence that methylation regulates termination *in vivo,* we have not pursued these observations further.}

The data of Figure 2 are for a member of the young PV Alu Subfamily, the APO Alu, which is cloned in a pUC vector (Materials and Methods). Virtually identical results have been obtained for seven different Alu subclones representing three distinct Alu subfamilies (Figure 1, Materials and Methods). One clone (Blur2) has no human DNA flanking sequences and in the case of another (EPL) removal of flanking sequences did not alter the effect of methylation. Also, similar effects were observed for Alu's cloned in either pUC or the TA cloning vector. These comparisons show that the effect of methylation does not depend critically on the nature of flanking sequences and can be attributed to methylation within the Alu template. Also, the effects of methylation are similar for a 172 base pair, left monomer, restriction fragment subcloned from EPL Alu, which like other human Alu's has a dimeric structure (Figure 1) . The EPL(L) construction, which includes the Pol III A and B boxes, deletes 5 CpG's located in the 3′ half of the dimeric Alu sequence. The remaining 13 CpG's located in the 5′ half of this truncated Alu are sufficient to cause repression (data not shown).

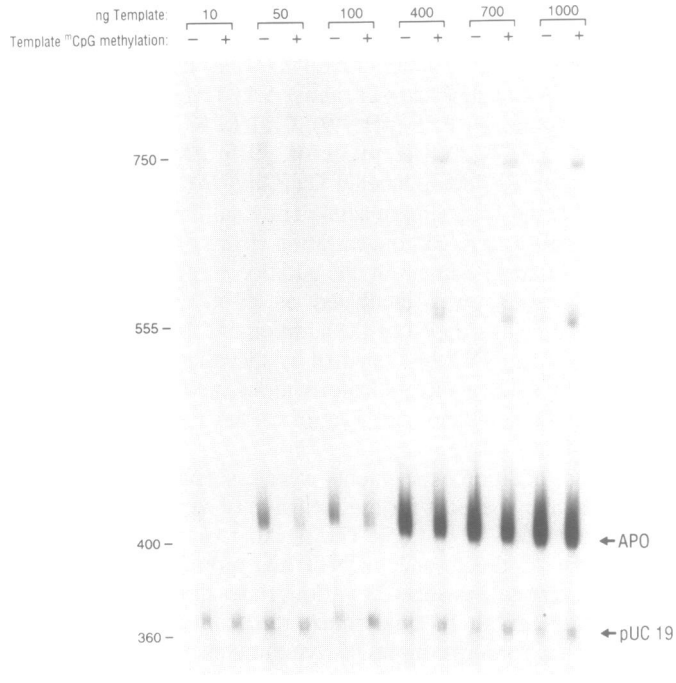



**Figure 2.** The template activity of CpG methylated (+) and unmethylated (−) APO Alu is examined at different concentrations. Total DNA is adjusted to one μg in each case by the addition of pUC19 DNA. Product lengths are given in nucleotides. The major product length designated by 'APO' corresponds to termination at the first run of four T's in flanking pUC sequence and additional minor product bands correspond to the lengths predicted for downstream termination in pUC sequence (Figure 1A). Vector products are designated by 'pUC'

**Table I.** Effect of template concentration on repression

| amount of template (ng) | 10 | 50 | 100 | 400 | 700 | 1000 |
|---|---|---|---|---|---|---|
| ratio of template activities | 2.7 | 2.8 | 1.28 | 1.20 | 1.22 | 1.07 |

The intensities of the major transcriptional products from the APO Alu template reported in Figure 2 were measured by radioanalytic imaging. The ratio is expressed as the intensity of the unmethylated template divided by the intensity of the methylated template corrected for background from a blank region of the gel.

## Methylation within or outside of the A box is sufficient for repression

The sequence CpGpCpG is fortuitously located only within the A box of both the EPL and APO PV Alu Subfamily members studied here (Figure 1B). Specifically, methylating this tetranucleotide with FnuDII DNA methylase represses transcription of both the APO Alu (not shown) and EPL Alu at low template concentration (Figure 3A). This effect must result from either the single FnuDII site in the A box or other sites within the plasmid vector. However, FnuDII methylation has no effect on the transcription of Blur2 which does not have any FnuDII sites within its Alu insert (data not shown). Therefore methylation of the vector DNA does not affect Alu transcription, while methylation of four C's located precisely in the A box is sufficient to cause repression.

To test whether the effect of methylation is limited to the A box, a mutant was derived from the EPL Alu in which all three CpG's within and near the A box are changed to CpA was constructed (EPL(A−), Figure 1). The 'A box mutant' Alu is transcribed, although at a greatly reduced level compared to its parent sequence (Figure 3B, see below). Methylation of this Alu repeat by CpG methylase outside of the A box represses its transcription relative to the unmethylated control (Figure 3B). Because of the very different template activities of EPL and EPL(A−), we cannot confidently compare possible differences in the levels of repression caused by methylating within and outside of the A box (Figure 3A,B). However, these findings and the previous results imply that methylation anywhere within or near the internal Pol III promoter of the Alu repeat can cause the repression.

## A repressor present in limiting amounts inhibits transcription

A DNA binding protein, MeCP1, is responsible for methylation sensitive repression of Pol II directed templates (23). This protein, present in limiting amounts, binds to any DNA sequence containing 5 me-C residues, although it prefers 5 me-C clusters (22). Competition by methylated pUC DNA relieves the transcriptional inhibition caused by methylating the template (Figure 4, Table II). In this experiment the 5 me-C composition of the pUC carrier is systematically varied while keeping total

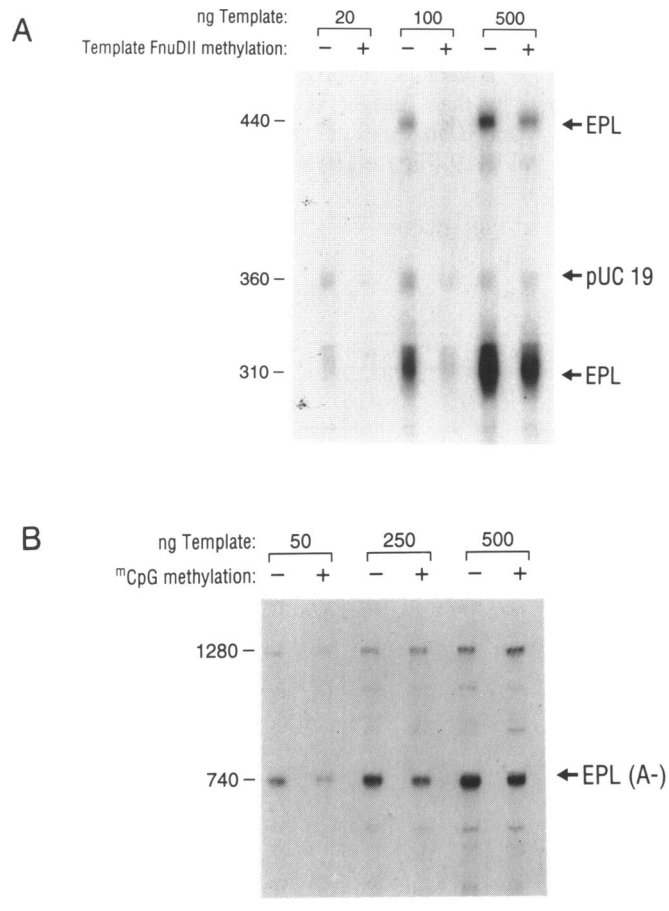



**Figure 3.** **(A)** The template activity of FnuDII methylated (+) and unmethylated (−) EPL Alu is examined at different concentrations. Total DNA is adjusted to a constant amount of one microgram in each case by the addition of pUC. The first terminator in this construct is leaky, giving rise to two major products, 'EPL', predicted for termination at two sets of four T's in the pUC sequence (Figure 1A). **(B)** Similarly, the template activity of CpG methylated and unmethylated EPL(A−) is examined at different concentrations.

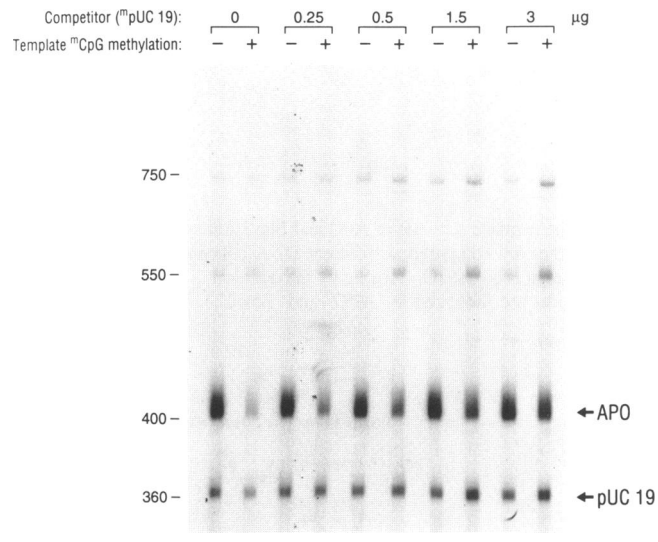



**Figure 4.** The template activity of CpG methylated (+) and unmethylated (−) APO Alu is examined at different concentrations of CpG methylated pUC DNA. The experiment is identical to that of Figure 2 except that 100 ng of template is used in each lane and total DNA is adjusted to three µg by the addition of methylated and unmethylated pUC DNA. The amount of methylated competitor is indicated above the figure.

**Table II.** Competition by methylated pUC DNA

| Methylated Competitor Composition | Activity of Unmethylated Template | Activity of Methylated Template | Ratio of Template Activities |
|---|---|---|---|
| 0% | 8.02 | 2.16 | 3.7 |
| 8.3% | 7.78 | 3.58 | 2.2 |
| 16.6% | 7.09 | 4.31 | 1.6 |
| 50.0% | 8.69 | 5.79 | 1.5 |
| 100% | 8.14 | 6.46 | 1.3 |

The intensities of the major transcriptional products at different compositions of methylated competitor DNA reported in Figure 4 were determined by radioanalytic imaging. These intensities are expressed in arbitrary phosphorescence units corrected for background from a blank region of the gel and reported above as template activity. The ratio of template activities is the activity of the unmethylated template divided by the activity of the methylated template.

DNA concentration constant. These data show that inhibition is caused by a repressor present in limiting amounts and that the repressor is not sequence specific but binds to exogenous methylated pUC DNA sequences. Of particular importance, the activity of the unmethylated template is constant at different concentrations of the methylated competitor showing both that the repressor is highly specific toward methylated DNA and that the competitor does not affect the activity of any other positive or negative transcription factors in these crude extracts (Figure 4, Table II). Methylated TA cloning vector DNA also competes for the repressor (data not shown). At the concentrations investigated here, methylated pUC did not completely relieve repression suggesting that the methylated Alu template may be a slightly more effective competitor than pUC (Tables I and II). The density of CpG's and their clustering may be important, as is the case for MeCP1 (22).

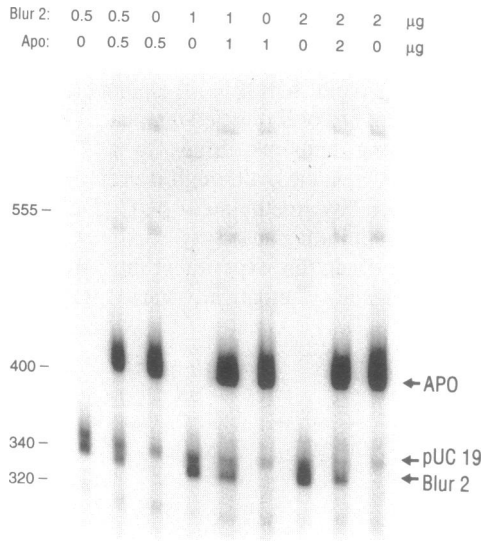



**Figure 5.** The template activities of APO and Blur2 Alu's alone and in competition with each other are compared at different DNA concentrations. Total DNA in each experiment is adjusted to four µg by the addition of pUC DNA. As before, major products are indicated as 'APO' and 'Blur2'.

MeCP1 is less abundant in mouse F9 embryonic cells than in HeLa cell extracts (22,23). When assayed at equal protein concentrations, the repression of methylated Alu templates is less pronounced in F9 than in HeLa extracts (data not shown). However, Pol III activity is also significantly lower in F9 cell extracts as compared to HeLa extracts (36, data not shown) so that we regard these observations as being inconclusive as to the identity of the methylated Alu repressor.

## Younger Alu's have a transcriptional advantage

Blur2 Alu was routinely transcribed less abundantly than any young Alu repeat investigated here. Depending on conditions, the APO Alu, a PV Subfamily representative, is transcribed about six times more abundantly than Blur2 (Figure 5). In direct competition assays, the APO Alu assumes an even greater transcriptional advantage, approaching an twelve-fold higher template activity than Blur2 (Figure 5). Because of its nearly consensus A and B boxes, Blur2 should be actively transcribed *in vitro* by Pol III (9, Figure 1B) so that this difference deserves attention.

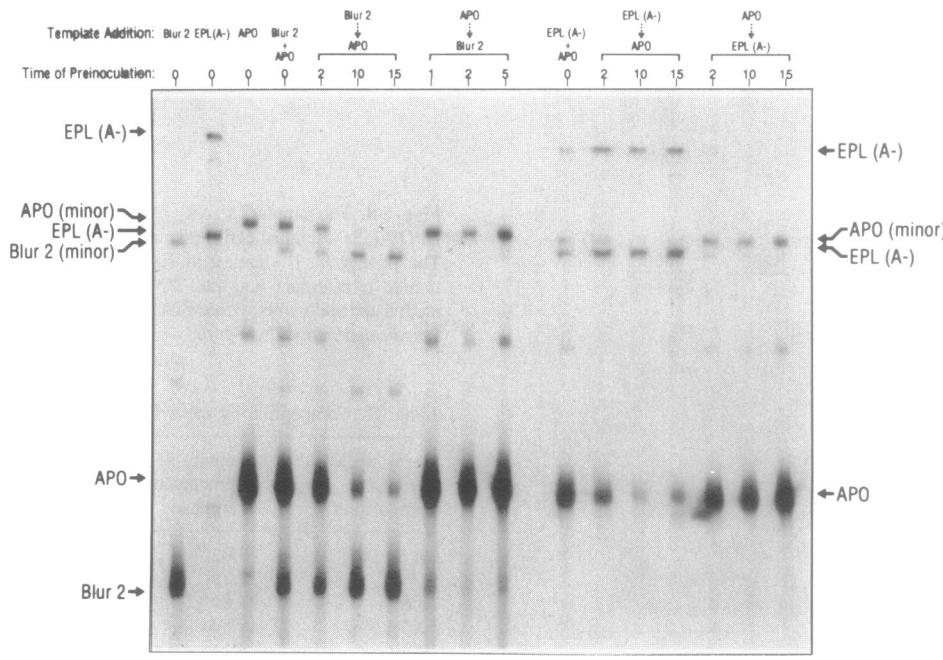



**Figure 6.** The effect of the order and time of addition on template activity is examined. In lanes having zero preincubation time, the templates are added directly to the extract and incubated for 15 minutes at room temperature before adding the nucleoside triphosphates at 30°C. These are Blur2, EPL (A−) and APO transcribed without competition, and both Blur2 in competition with APO, and EPL (A−) in competition with APO. In the other lanes, one template is preincubated with the extract for the indicated period of time before adding the second template. One µg of each template is used giving a total of two µg of DNA in the competition assay and pUC is added to adjust total DNA to two µg in the three assays without competition.

There is of course considerable heterogeneity in the template activity of Alu repeats. Maniatis and Shen compared many Alu repeats resident in the human $\alpha$- and $\beta$- globin gene clusters (5,6). Of these Alu's, the most active by far is located 3′ to the human $\alpha$-1 hemoglobin gene, the 3′$\alpha$-1 Alu (5,6). We find that the 3′$\alpha$-1 Alu, which happens to have perfect young Alu subfamily A and B boxes (Discussion), has 75% of the template activity of the Apo Alu (data not shown) . The template activity of Blur2, its CpG content, and its percentage divergence from its subfamily consensus sequence are typical of most other members of the Major Alu Subfamily, which are old with respect to the evolutionary time of their appearance in the ancestral human genome. For these reasons, we regard Blur2 as being representative of older Alu's.

Why is the APO Alu more abundantly transcribed than Blur2? This question is partially answered by the mechanism of Pol III transcription which requires recognition of a promoter, formation of a stable complex, and subsequently, multiple rounds of transcription (19,34,35,37). Order of addition experiments classically distinguish between these steps; a template added first kinetically competes for any limiting factors at the expense of a second template added at later times (19,37). Blur2 is the more active template when it is added first to the extract and followed by the APO Alu (Figure 6). The kinetics of this process are potentially revealing. The first order rate constant for forming a stable transcription complex with Blur2 corresponds to a reaction life time of nine minutes whereas the life time of forming a stable complex with the APO Alu is less than one minute (Figure 6). The relatively rapid rate of formation of the Pol III stable complex with the APO Alu template is sufficient to explain its transcriptional advantage in competition with Blur2.

Formation of the stable complex is thought to processed by interaction of TFIII C with the promoter A and B boxes followed by an interaction of this complex with TFIII B (19,37). The mutant A box Alu, EPL(A−), is also significantly less active than the APO Alu template (Figure 6). The transcriptional activity of the EPL Alu parent clone relative to the mutant A box Alu is entirely comparable to the activity of the APO Alu template (Figure 3A and data not shown), so that converting the three CpG's in and near the A box to CpA's dramatically reduces template activity. However, despite this reduction in activity, these three base substitutions still conform to the A box consensus sequence (Figure 1B). In order of addition experiments, EPL(A−),like Blur2, also forms stable complex more slowly than the APO Alu (Figure 6) while the EPL Alu parent to the mutated A box template competes for formation of the stable complex at the same rapid rate as the APO Alu (data not shown). These three targeted CpG mutations within the A box of a young subfamily Alu repeat cripple its kinetic competition for limiting factors and its template activity.

Both Blur2 and the mutant A box Alu have lower template activities than the APO Alu. However the reduced transcriptional activity of Blur2 could result from mutations in either its A box or elsewhere within its sequence (Figure 1B). To rescue Blur2's template activity, its A box was corrected to exactly match the A box of young PV subfamily Alu repeats including the three CpG's (Figure 1B). This correction slightly improves the template activity of Blur2 (ranging from 110% to 160% of Blur2's activity, data not shown). Also in order of addition experiments, both the corrected and uncorrected Blur2 templates form stable complexes at similarly slow rates compared to the APO Alu (data not shown). Evidently, a perfect A box having all three CpG's is necessary but not sufficient for both rapid formation of the stable complex and high template activity. Blur2 has a consensus B box which however differs from the PV consensus sequence at two positions (Figure 1B).

## DISCUSSION

### Methylation indirectly represses Alu transcription

Previously, the effect of methylation on the *in vitro* transcription of Alu repeats was not detectable (5). Here we used low rather than high template concentrations, nuclear rather than cytoplasmic extracts, very CpG rich young Alu templates rather than a single older Alu, and either general methylation by CpG methylase or methylation targeted specifically to the A box rather than the untargeted methylation of a few restriction sites. While each of these technical differences undoubtedly contributes somewhat to the detection of methylation induced repression, the use of low template concentrations is critical (21).

In exact analogy to the VAI RNA gene, methylation inhibits the *in vitro* transcription of Alu repeats at low but not at high template concentrations (21). While we cannot exactly compare their levels of inhibition, results obtained for the VAI gene are similar to those observed here for Alu, which is also a Pol III type 2 template (Introduction). In principle, this inhibition could result from either limiting amounts of a repressor that is specific to methylated templates or by a lower affinity of positive transcription factors toward methylated templates. By a gel mobility shift assay, methylation of the VAI RNA gene interferes with formation of a protein DNA complex leading to the hypothesis that methylation inhibits template activity by interfering with the interaction of positive transcription factors (21). However, the inhibition of Alu templates is relieved by competition with methylated plasmid DNA, showing that it results from repression. Also, methylated competitors have no effect on the activity of unmethylated templates, showing that the repressor is highly specific for methylated templates. While specific toward 5 me-C, this factor does not bind to a specific sequence. One possibility not examined here is whether this factor, like MeCP1, prefers clusters of 5 me-C's (22). Circumstantial evidence suggests that the 5 me-C Alu repressor is probably MeCP1.

While we have no direct evidence that this factor represses Alu transcription *in vivo,* methylation inhibits the *in vivo* transcription of other type 2 templates (tRNA genes and the VAI RNA gene) so that a similar effect on CpG rich Alu repeats is likely. However, the transcription of Alu members is probably regulated at many levels and might depend on its chromatin context, presence of cis acting elements and other factors in addition to its methylation status (14). For example, a stably transfected PV Alu repeat is transcriptionally silent in mouse cells whether introduced as a pUC subclone or even in the active chromatin context of an episomal BPV vector (16). These transfected sequences remained unmethylated showing that methylation is not necessary for repressing Alu transcription. The global methylation of Alu repeats in many somatic tissues possibly 'locks' out the inappropriate expression of individual members of this very large repeated sequence family (38).

### Transcriptional advantage of young Alu members

5me-C is essentially a potent endogenous mutagen that rapidly causes C to T transitions (39). In contrast to older Alu's, young Alu's, ie. those which recently appeared in the ancestral human

genome, have almost all CpG's intact; the conservation of these residues must result from either selection, correction or maintenance (Introduction). As one of several possibilities, differences in template activity might give young Alu's a transcriptional and consequent transpositional advantage over older Alu repeats.

Here, a representative older Alu, even with corrected consensus A and B boxes, is both less active and forms a stable transcription complex slower than a young Alu template. In effect, this representative older Alu is selectively disadvantaged at two distinct steps in the transcriptional pathway: In competition for limiting transcription factors *in vivo*, the younger Alu template has a kinetic advantage. And, upon forming a stable complex, younger Alus are apparently more abundantly transcribed. The multiplicative advantage provided by these two steps could greatly favor the *in vivo* transcription, and consequently the transposition, of younger Alu's.

The A box is especially important in determining template activity (6,9,40). Specifically, mutating the three CpG's to CpA's within and near the A box without disrupting its consensus sequence, drastically reduces the template activity of a young Alu repeat, showing that at least some of these three CpG's associated with the A box are important. Order of addition experiments further show that these same mutations greatly decrease the rate of stable complex formation. Again, as discussed above, the multiplicative effects of reducing both template activity and the rate of competition for potentially limiting factors place an Alu with these mutations at an extreme disadvantage compared to Alu's with intact A box CpG's. Correcting the A box of the Blur2 Alu is not sufficient to completely restore either its template activity or its rate of stable complex formation, so that mutations outside the A box are also important.

Considerable heterogeneity in the template activity of Alu sequences is well documented (4). However, neither this nor previous investigations have attempted to define precisely the minimum number of mutations or combinations of mutations which inactivate Alu members. In one comparison of many Alu's interspersed within the human $\alpha$- and $\beta$- hemoglobin gene clusters, Shen and Maniatis observed a spectrum of template activities, but the most active Alu by far resided 3′ to the $\alpha$- globin gene cluster (4). Interestingly, of the many Major Subfamily Alu's surveyed in that study, the 3′ $\alpha$-globin Alu is the only one having an A box and a B box that exactly matches the PV Subfamily consensus sequence. Thus, while we are unable to define the most important mutations inactivating Alu templates, the very best templates *in vitro*, without exception, exactly match the PV and Precise Subfamily consensus A and B boxes. Flanking sequences affect the activity of many Pol III directed templates (41–43). The possible effect of flanking sequences on Alu templates is intentionally not investigated here and our conclusions concern only their internal Pol III promoter.

All Alu subfamilies are transcribed *in vivo* by Pol III in HeLa cells, although there is a difference of opinion as to whether young Alu subfamily members are overrepresented in these transcripts (10–12). Of fifteen cDNA's representing Pol III Alu transcripts from HeLa cells, ten have all three A box CpG's intact (12). Based on Jurka and Milosavljevic's (44) sequence data base analysis, only 7% of all Alu members would have these three CpG's intact. Also, of nine Alu's that are young in the sense of not being fixed in the human population, all nine have intact CpG's in the A box. This direct biological evidence for the importance of these three CpG's in Alu transcription and transposition complements the biochemical evidence for the importance of these CpG's summarized above. The A box is an especially critical determinant of template activity (6,9,40). Transcriptional and transpositional activity are certainly regulated at several distinct steps, but 5 me-C to T transitions at critical sites in the A box are sufficient to place older Alu's at a selective disadvantage.

In summary, these results and additional considerations presented in the Introduction and Discussion lead us to two intriguing interlocked hypotheses: Older Alu's are not especially active templates due to accumulated mutations, especially the CpG transitions, but the younger potentially most active Alu templates are subject to repression by methylation at many of the same critical CpG sites.

## ACKNOWLEDGEMENTS

We greatly appreciate helpful discussions with Drs Carol Rubin, James Shen, Michael Dahmus, Gregg Morin and Prabhakara Choudary, who also graciously supplied us with F9 cells. We also thank Dr Igor Chesnokov for providing the EPL parent clone. This research was supported by USPHS Grant GM 21346 and the Agricultural Experiment Station of the University of California.

## REFERENCES

1. Weiner, A.M., Deininger, P.L. and Efstratiadis, A. (1986) *Annu. Rev. Biochem.* **55**, 631–661.
2. Schmid, C.W., Deka, N. and Matera, A.G. (1990) In Adolpf, K.W.(ed) Chromosomes eukaryotic prokaryotic and viral. CRC Press, Boca Raton Florida, pp. 323–358.
3. Fuhrman, S.A., Deininger, P.L., LaPorte, P., Friedmann, T. and Geiduschek, E.P. (1981) **9**, 6439–6456.
4. Shen, C.-K.J. and Maniatis, T. (1982) *J. Mol. Appl. Gen.* **1**, 343–360.
5. Shen, C.-K.J. and Maniatis, T. (1982) In Wu, H.-K. (ed) Genetic Engineering Techniques. Academic Press, N.Y., pp.129–158.
6. Perez-Stable, C., Ayres, T.M. and Shen, C.-K.J. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 5291–5295.
7. Segni, G.D., Carrara, G. and Tocchini-Valentini, G.R. (1981) *Nucleic Acids Res.* **9**, 6709–6722.
8. Liu, W.-M., Leeflang, E.P. and Schmid, C.W. (1992) *Biochim. Biophys. Acta* **1132**, 306–308.
9. Paolella, G., Lucero, M.A., Murphy, M.H. and Baralle, F.E. (1983) *EMBO J.* **2**, 691–696.
10. Matera, A.G., Hellmann-Blumberg, U. and Schmid, C.W. (1990) *Mol. Cell Biol.* **10**, 5424–5432.
11. Sinnett, D., Richer, C., Deragon, J.M. and Labuda, D. (1992) *J. Mol. Biol.* **226**, 689–706.
12. Maraia, R.J., Driscoll, C., Bilyeu T., Hsu, K., and Darlington, G.J. (1993) (submitted).
13. Deininger, P.L., Batzer, M.A., Hutchison III, C.A. and Edgell, M.H. (1992) *Trends in Genet.* **8**, 307–311.
14. Schmid, C.W. and Maraia, R. (1992) *Current Opinion in Genetics and Development: Genomes and Evolution* **2**, in press.
15. Shen, M.R., Batzer, M.A. and Deininger P.L. (1991) *J.Mol. Evol.* **33**, 311–320.
16. Leeflang, E.P., Liu, W.-M., Hashimoto, C., Prabhakara, C.V. and Schmid, C.W. (1992) *J. Mol. Evol.* **35**, 7–16.
17. Schmid, C.W. (1991) *Nucleic Acids Res.* **19**, 5613–5617.
18. Hellmann-Blumberg, U., Hintz, M.F. and Schmid C.W. *Mol. Cell Biol.* (submitted).
19. Geiduschek, E.P. and Tocchini-Valentini, G.P. (1988) *Ann. Rev. Biochem.* **57**, 873–914.
20. Besser, D., Gotz, F., Schulze-Forster, K., Wagner, H. , Kroger H. and Simon, D. (1990) *FEBS Let.* **269**, 358–362.
21. Juttermann, R., Hosokawa, K. Kochanek, S. and Doerfler, W. (1991) *J. Virol.* **65**, 1735–1742.
22. Meehan, R.R., Lewis, J.D., McKay, S., Kleiner, E.L. and Bird, A.P. (1989) *Cell* **58**, 499–507.

23. Boyes, J. and Bird, A. (1991) *Cell* **64**, 1123–1134.
24. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor University Press, Cold Spring Harbor.
25. Dignam, J.D., Lebovitz, R.M. and Roeder, R.G. (1983) *Nucleic Acid Res.* **11**, 1475–1489.
26. Bradford, M.M. (1976) *Analy. Biochem.* **72**, 248–254.
27. Deininger , P.L., Jolly, D.J., Rubin, C.M., Friedmann, T. and Schmid, C.W. (1981) *J. Mol. Biol.* **151**, 17–33.
28. Mietus-Snyder, M., Charmley, P., Korf, B., Ladias, J.A.A., Gatti, R.A. and Karathanasis, S.K. (1990) *Genomics* **7**, 63–67.
29. Matera, A.G., Hellmann-Blumberg, U., Hintz, M.F. and Schmid, C.W. (1990) *Nucleic Acids Res.* **18**, 6019–6023.
30. Leeflang, E.P., Liu, W.-M., Chesnokov, I. and Schmid C.W. *J. Mol. Evol.* (submitted).
31. Gibbs, P.E., Zeilinski, R., Boyd, C. and Dugaiczyk, A. (1987) *Biochemistry* **26**, 1332–1343.
32. Ryan, S.C. and Dugaiczyk, A. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 9360–9364.
33. Platt, T. (1986) *Ann. Rev. Biochem.* **55**, 339–372.
34. Lassar, A.B., Martin, P.L. and Roeder, R.G. (1983) *Science* **222**, 740–748.
35. Schaack, J., Sharp, S., Dingermann, T. and Soll, D. (1983) *J. Biol. Chem.* **258**, 2447–2453.
36. White, R.J., Stott, D. and Rigby, P.W.J. (1989) *Cell* **59**, 1081–1092.
37. Palmer, J.M. and Folk, W.R. (1990) *TIBS* **15**, 300–304.
38. Selker, E.U. (1990) *TIBS* **15**, 103–107
39. Rideout III, W.M., Coetzee, G.A., Olumi, A.F. and Jones, P.A. (1990) *Science* **249**, 1288–1290.
40. Hofstetter, H., Kressmann, A. and Birnstiel, M. (1981) *Cell* **24**, 573–585,
41. Ullu, E. and Weiner, A.M. (1985) *Nature* **318**, 371–374.
42. Murphy, S., Di Liegro, C. and Melli, M. (1987) Cell **51**, 81–87.
43. Lobo, S.M. and Hernandez, N. (1989) *Cell* **58**,55–57.
44. Jurka, J. and Milosavljevic, A. (1992) *J. Mol. Evol.* **32**, 105–121.