Nucleic Acids Research, Vol. 20, No. 12 2941–2945

Differential binding of human nuclear proteins to Alu subfamilies

N.V.Tomilin, V.M.Bozhkov, E.M.Bradbury[1] and C.W.Schmid[2]*
Institute of Cytology, The Academy of Sciences of the USSR, Leningrad 194064, Russia,
[1]Department of Biological Chemistry, University of California, Davis, CA 95616 and [2]Departments of Genetics and Chemistry, University of California, Davis, CA 95616, USA

Received April 22, 1992; Accepted May 12, 1992

ABSTRACT

Several diagnostic differences that distinguish human Alu subfamilies are clustered just downstream from the B box of the RNA polymerase III promoter; we tentatively refer to this diagnostic region as the DB box. Assuming that this region might determine the relative transcriptional activity of Alu subfamilies, we examined the interaction of nuclear proteins with DB box sequences representing different Alu subfamilies. Gel mobility shift assays suggest the existence of two factors which discriminate among the DB boxes of different Alu subfamilies: 1) An abundant, ca. 50 kd, protein binds more stably to a young 'PV' Alu subfamily (PVS) than to the older major subfamily (MS). 2) Methylation of CpG dinucleotides stimulates the binding of a less abundant, ca. 70 kd, protein to the DB boxes of younger Alu subfamilies.

INTRODUCTION

By sequence analyses several groups have concluded that human Alu repeats can be divided into recognizable subfamilies that differ in their evolutionary time of appearance within the human genome (1–6). This conclusion has been confirmed and extended by the hybridization of probes targeted toward diagnostic sequence features of the putative youngest Alu subfamily (7–10). The youngest, so called PV, subfamily (PVS) recently expanded in the human lineage following the divergence of human and apes, includes members which transposed in the contemporary human population and includes one or more transcriptionally active locus (7–10). The precise subfamily (PS) is evolutionarily older but may include both transcriptionally and transpositionally active loci (7). Members of the older and larger major subfamily (MS) are fixed in the human lineage over an evolutionarily long period of time and are no longer transcriptionally or transpositionally active (1,2,7,9). [Nomenclature is unresolved and while here we use the subfamily names introduced by this laboratory it should be noted that essentially identical Alu subfamilies have been identified by the Deininger group using different names (9). In particular, the Deininger group HS subfamily is identical to the PV subfamily discussed here].

The existence of these Alu subfamilies and their different transcriptional activities raises the issue of how these differences are determined. Whereas most Alu repeats contain an internal promoter for RNA polymerase III and are transcribed *in vitro*, most Alu members are transcriptionally silent in HeLa cells (11–13). Among other possibilities, the diagnostic sequence differences of the different Alu subfamilies might also account for their different transcriptional activity. The PVS and MS consensus sequences differ at many sites scattered through the entire Alu repeat which are depicted in Table I for positions 76 to 107. Surprisingly, there are only a few differences in the left end region (A and B boxes of RNA polymerase III promoter), known to be important for *in vitro* Alu transcription. The GC rich space between the A and B boxes is the same in the PVS and MS sequences. This same region interacts with a human zinc-finger nuclear protein which has been suggested to be a repressor of Alu transcription *in vivo* (14, 15, 28) However, a cluster of five differences between PVS and MS is located in the region (named here the DB region) just downstream from the B box of the Alu RNA polymerase III promoter (Table I). Interestingly, this region resembles the A/T box of the U6 snRNA gene promoter (Table I). The U6 A/T box is known to be a dominant cis-element of RNA polymerase III transcription of U6 genes (16,17).

The PVS and PS sequences differ by two nucleotides in the DB region (Table I comparing the PV and PR probes). If the DB region is important for the control of Alu transcription *in vivo*, it should interact with positive and/or negative transacting-factors and the stability of this interaction should be related to the origin of the DB region e.g. PVS versus PS versus MC. Using the gel mobility shift assay, these hypotheses are tested with a set of variant synthetic oligonucleotides representing these different Alu subfamilies.

MATERIALS AND METHODS

Alu-DB region oligonucleotides

Double strand oligonucleotides representing various Alu subfamilies or specific sequence variants were prepared by the hybridization of oligonucleotide complements and filling in the

* To whom correspondence should be addressed

resultant hybrid by DNA polymerase. The various single strand oligonucleotides used in these constructions have ten to fourteen overlapping complementary nucleotides. In each case the upper strand as depicted in Table I carried the radioactive label. The oligonucleotide complements used to form the substrates shown in Table I are as follows

PV:	ATCGAGACCATCCCGGCTAAAA	plus GGTTTCACCGTTTTAGCCG
PV51:	ACCATCCCGGCTAAAACGGTGA	plus GGTTTCACCGTTTTAGCCG
PR:	ATCGAGACCATCCTGGCTAACAC	plus GGTTTCACCGTGTTAGCCA
PRM:	ATCGAGACCATCCTGGCTAAAACG	plus GGTTTCACCGTTTTAGCCAG
PRC:	ATCGAGACCATCCCGGCTAACACG	plus GGTTTCACCGTGTTAGCCGGG
PVU:	ATCGAGACCATCCCGGCTTTATATATG	plus GGTTTCATATATAAAGCCGGG
MC:	TTCGAGACCAGCCTGGCCAACAT	plus GGTTTCACCATGTTGGCCA
Blur 2:	TTCAAGACCAGCTTGACCAACAT	plus GGTTTCACCATGTTGGTCA

The sequence differences between the resulting duplex products are depicted in Table I.

To form the duplex products, the oligonucleotides identified above were subjected to the following procedure: 200 ng of the upper strand oligo was 5′ labeled with ^{32}P in 20 μl reaction mix containing 6 μl (60 μCi) [γ-^{32}P]-ATP (Amersham, sp. act. 3000 Ci/mmol), 2 μl of 10×kinase buffer and 10 units of T4 polynucleotide kinase (Boehringer). After 1 hour at 37°C, the enzyme was inactivated by heating for 5 min. at 80°C. 8 μl of non-labeled dNTP mix (2 nM each), 3 μl of 10×nick translation buffer, an equimolar amount of corresponding non-labeled lower-strand oligo and 5 units of the Klenow DNA polymerase (Boehringer) were added and the final volume was adjusted to 50 μl with water. After incubation for 1 hour at 25°C, 2 μl of 0.5 M EDTA and equal volume of phenol-chloroform were added followed by 5 min. of shaking and 5 min. of centrifugation at 25°C. The aqueous phase was applied to a G25 Sephadex column (1.5 ml) and eluted with TE buffer. The exclusion peak of radioactivity (100–200 ml) was used as substrate in GMSA.

Aliquots of these duplex substrates run on a 4% polyacrylamide gel were examined in the absence of nuclear extracts (Fig. 1A). In each case there is a single major band having mobility slightly higher than bromophenol blue dye. As expected, the band obtained for the PV51 product showed a higher mobility (Fig. 1, lane 1) than the other substrates. Examination of some of the substrates on a 12% sequencing gel along with chemically synthesized 32-mer (labeled at 5′ end using polynucleotide kinase) as a marker showed that more than 80% of the radioactive strands in the GMSA substrates are 32-mers. By the same assay, the PV51 substrate mainly contain, as expected, radioactive 26-mers.

Table I.

	76	82	89	96	117
PV:	ATCGAGACCATCCCGGCTAAAACGGTGAAACC				
PV51:	XXXXX----------				
PR:	----------T------C----------				
PRM:	----------T----------				
PRC:	----------C----------				
PVU:	----------TTT-TATAT-----				
MC:	T---------G--T---C--C-T---------				
Blur2:	T--A------G-TT-A-C--C-T---------				

Upper strand sequences of the DB regions of the different Alu subfamilies and sequence variations in this investigation are compared. A dash (−) indicates the same sequence, base substitutions are indicated by the appropriate letter and X indicates a deletion. PV and PV51 match the PV subfamily consensus, PR matches the Precise subfamily consensus and MC the Major subfamily consensus sequences in the regions depicted (2, 7, 8). Blur2 can be regarded as a particular member of the Major subfamily. PRM, PRC and PVU test specific sequence changes as described in the text.

Gel mobility shift assay

Gel mobility shift assays (GMSA (18, 19)) were performed as described previously (15) and in the legend to Fig. 2A. In most

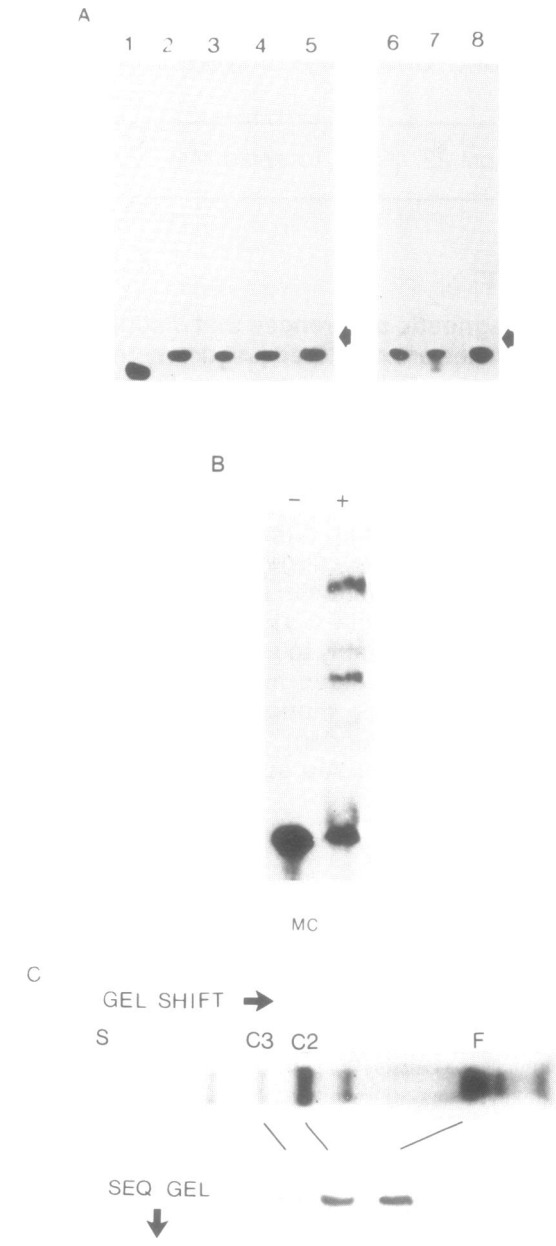

Figure 1. Properties of the GMSA substrates. (A) The mobilities of the substrates are compared by polyacrylamide gel electrophoresis as described in Materials and Methods. Lanes 1 through 8 are PV51, PVU, Blur2, PR, PRM, MC, PV and PR, respectively. The results of different runs (lanes 1–5 and lanes 6–8) are shown. Filled arrows show the mobility of the bromphenol blue dye. (B) GMSA with MC substrate in the presence (+) or in the absence (−) of HeLa nuclear extract with 50 ng of the non-labeled competitor salmon DNA. GMSA was performed as described in the legend to Fig. 2 and Materials and Methods. (C) Analysis of the GMSA bands (Blur2 substrate) on a 12% sequencing gel. GMSA was run as in (B), the bands were eluted and following nucleic acid extraction run on the second denaturing gel as indicated.

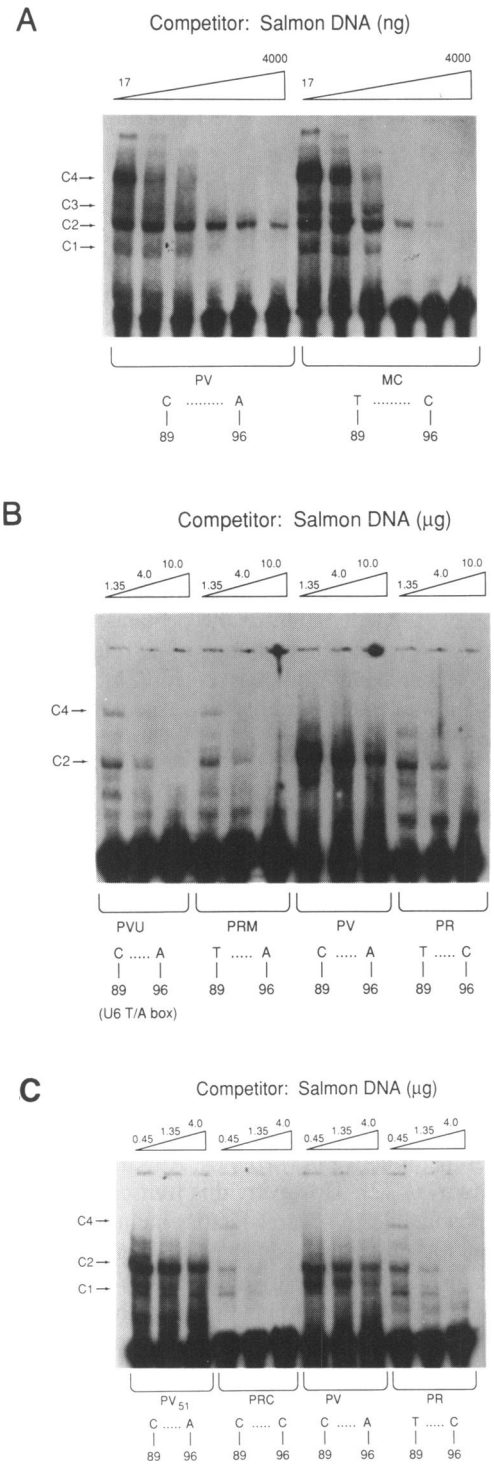

Figure 2. Gel mobility shift assay of the binding of HeLa nuclear proteins to the oligonucleotides covering the region of Alu repeats just downstream from the B box of the RNA polymerase III promoter (Table I). Panel A—complexes detected with PV and MC (major subfamily consensus) oligonucleotides with different concentrations of non-specific competitor salmon DNA. C1–C4 indicate complexes with proteins. Incubation mixes (20 μL) contained 4 ng of oligonucleotide with the upper strand labeled with 32-P using T4 polynucleotide kinase, 40 mM KCl, 1 mM EDTA, pH 8, 1 mM dithiothreitol, 1 mM Tris-HCl, pH 8, 2 mM HEPES, pH 8, an indicated amount of non-labeled competitor DNA, 2 μL of HeLa nuclear extract. After incubation for 30 min. at room temperature samples were run in a 4% polyacrylamide gel (5% bis) in 0.25×TBE, dried and autoradiographed. Panel B—binding at high stringency of PVU, PRM, PV and PR oligos; C—binding of PV, PV_{51}, PRC and PRM oligos, experimental conditions are the same as in A.

experiments 40 mM KCl was present in the binding mix except the experiment shown in Fig. 3 where the KCl concentration was 120 mM. Nuclear extract was prepared from HeLa cells by microscale extraction of isolated nuclei with a buffer containing 0.4 M NaCl (15).

Methylated PV and PR oligonucleotides were prepared using prokaryotic CpG methylase M.SssI (New England Biolabs) in a buffer provided by the company, and before GMSA were deproteinized with phenol-chlorophorm, purified using a micro-column of Sephadex G2ecipitated with ethanol. The variants used as controls in GMSA were processed in the same way but in the absence of M.SssI. The efficiency of methylation was confirmed by the appearance of an HpaII-resistant site in PV substrate containing CCGG sequences (Table 1).

RESULTS

Additional GMSA Controls

As shown below, a number of complexes are detected by GMSA. These apparent complexes do not form in the absence of nuclear extract but depend on the addition of the extract (Fig. 1B). The mobility difference of these putative complexes might be attributable to either different nucleoprotein particles or some inherent unsuspected heterogeneity in the labeled substrate. To test this possibility, the labeled substrate was extracted from two of the principle complexes, which are called C2 and C3 (see below), and compared to the free uncomplexed substrate on a denaturing sequencing gel (Fig. 1C). The mobilities of the oligonucleotides isolated from the C2 and C3 complexes are indistinguishable from the mobility of the uncomplexed substrate.

Several complexes form with the region downstream from the B box 'DB box'

By GMSA, the DB boxes of Alu subfamilies form several complexes with HeLa nuclear proteins (Fig. 2). At least four such complexes (named C1 through C4) are apparent for the oligonucleotides that match the major family consensus (MC, Fig. 2 A). Confirming this result, an almost identical pattern (data not shown) was observed using an oligonucleotide that matches the sequence of a representative example of the major subfamily, Blur 2 (Table I).

We assume that the observed complexes actually are DNA-protein complexes (18, 19). To estimate the sizes of proteins

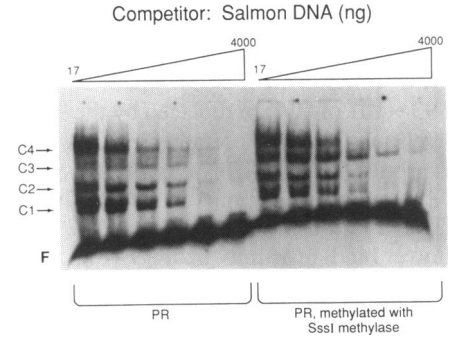

Figure 3. The influence of methylation of PR oligonucleotide on the formation of different complexes with HeLa proteins. Experimental conditions were the same as in Fig. 2A except 120 mM KCl was present in binding mixes.

forming these complexes, we further assume (21) that the relative mobility (RM) depends linearly on the logarithm of a protein molecular weight for complexes with short oligonucleotides and that RM = 0.3 corresponds to M = 100 kd (21). Estimated in this way approximate molecular masses of proteins forming C4, C3, C2 and C1 were found to be 100 kd, 70 kd, 50 kd and 30 kd, respectively. Dilutions of the nuclear extract (data not shown) suggest that the C1 factor is the most abundant. However, C2 is also readily detected even upon an eighty fold dilution. The abundance and low molecular weight of C1 suggests it might result from an HMG protein. Another issue is whether the stabilities of these complexes depend on the sequence differences which distinguish Alu subfamilies. As one example, the stability of C1 does not noticeably discriminate between different Alu subfamilies (Fig. 2, A, B,C, Fig. 3).

Dependence of C2's stability on Alu subfamilies

In contrast, the complex C2 (RM = 0.5) is more stable with PV oligonucleotides than with representatives of other subfamilies (Fig. 2 A, B, C). Since oligonucleotides PV and MC differ at six sites, it is perhaps not especially surprising that their corresponding C2 complexes have different stabilities (Fig. 2A). However, C2 also forms a more stable complex with PV than with PR probes (Fig. 2B, 2C). Since oligonucleotides PV and PR differ at only two sites (position 89 and 96), this region must be involved in formation of C2 (Table I, Fig. 2 B, 2C). Confirming the importance of this region, representative PV probes were constructed using two different upper strands, PV_{51} and PV (Table I). In each case, the C2 complex formed with PV probes was more stable than the complexes formed with probes representing other subfamilies (Fig. 2B, 2C). [These results also suggest that the presence of the minor labeled contamination in the PV (Fig. 1A, lane 7) as compared to PV51 (Fig. 1A, lane 1) does not significantly influence the results obtained in GMSA].

Either the C at position 89 or the A at position 96 or the combination of C and A at positions 89 and 96 must account for the different stability of C2 with the PV and closely related PR sequences. To distinguish these alternatives, all combinations of changes were compared (Table I). Substitution of T for C at position 89 (PRM Fig, 2B) decreases C2 stability (Fig. 2B) and similarly substitution of C for A at position 96 (PRC, FiC) also decreases C2 stability. Since the stability of C2 with each of these single nucleotide 'mutations' approximates that of C2 with PR, we conclude that both positions 89 and 96 are probably important for complex formation.

Interestingly, the DB region of Alu resembles the A/T box promoter region of the human U6 gene (16,17). To examine whether the A/T box sequence is important for C2 formation, we used PVU oligo which is identical to PV except that nucleotides 93–101 of PV are replaced by the 9-nucleotide A/T box of U6 gene which is important for RNA polymerase III-driven transcription (16,17). This oligo has the same critical C at position 89 and A at position 96 as does the PV probe but A96 is embedded into the context of the U6 A/T box (Table I). PVU has a lower binding affinity than PV (Fig. 2B, lanes 1–3), suggesting that the nucleotide context around the critical C89 and A96 positions is also important for high-affinity C2 binding. This finding, namely that other sites are important, is consistent with observations (not shown) that the major family oligo MC which differs at four positions (76, 86, 93, 98) from PR has a lower binding affinity than PR.

CpG methylation stimulates C3 formation

Members of the young Alu subfamilies are especially rich in CpG dinucleotides (7,9) and are extensively methylated *in vivo* (22). Methylation of these sites might affect protein-DNA interaction, a possibility examined below.

At 40 mM KCl (Fig. 1A), C3 preferentially forms with the MC oligonucleotide probe relative to the PV oligonucleotide. In additional experiments, we observed that whereas higher salt (120 mM KCl) suppresses formation of C1, C2 and C4, higher salt stimulates the formation of C3 with the PR oligonucleotide. To enhance C3 formation, the following study was conducted at higher salt. Methylation of CpGs by SssI methylase markedly stimulates C3 formation for both PR (Fig. 3) and PV oligonucleotides (not shown). Since these oligonucleotides have only a single common CpG sequence near the region important for protein binding (Table I), we conclude that the region around position 95 to 101 is probably involved in C3 formation.

Interestingly, this region in the PR sequence (CACGGTG) contains an inverted repeat resembling the binding site for the c-myc protein (CACGTG) which is suppressed by CpG methylation (23). Methylation of the pentanucleotide CACGG is thought to stimulate the binding of a relatively abundant methylation sensitive DNA binding protein (24). Because of C3's abundance and molecular weight, we suspect that this complex might involve this same protein.

DISCUSSION

Because of their abundance Alu repeats may be a major determinant of chromosome structure. Consistent with this suggestion Alus are preferentially concentrated within R bands of human metaphase chromosomes (27). A structural role for Alu repeats could presumably involve their sequence specific interaction with relatively abundant proteins. Perhaps coincidentally, one subunit of human chromatin assembly factor (CAF-I) has a molecular weight (50 kd) similar to that of the C2 protein described here (29).

Although Alu repeats typically contain an RNA polymerase III promoter that is active *in vitro* most Alus are transcriptionally silent *in vivo (11, 12)*. However, this transcriptional inactivity might be attributable to several factors including an additional requirement for upstream sequences and an active chromatin context, or the extensive methylation of Alu repeats (reviewed in 22). Alternatively, one might imagine that nuclear factors binding in a sequence specific manner discriminate between Alu subfamilies thereby either silencing the majority or activating the minority of Alu. Although we have not tested the functional significance of the complexes reported here, the following empirical observations are noteworthy: Alu repeats complex with relatively abundant nuclear factors suggesting a possible role in chromosome structure. Further, the binding of certain factors is exquisitely sensitive to the minor sequence differences that distinguish the younger recently active Alu repeats from members of the older subfamily; plausibly these same factors might account for the different activities of these subfamilies.

ACKNOWLEDGEMENTS

This research was supported in part by USPHS grants GM-21346 (CWS) and GM 26901 (EMB). The authors thank Ms. U.Hellmann for providing several oligonucleotide samples.

REFERENCES

1. Slagel, V., Flemington, E., Traina-Dorge, V., Bradshaw, H., Jr. and Deininger, P.L. (1987) *Mol Biol. Evol.* **4**, 19–29.
2. Willard, C., Nguyen, H.T. and Schmid, C.W. (1987) *J. Mol. Evol.* **26**, 180–186.
3. Britten, R.J., Baron, W.F., Stout, D.B. and Davidson, E.H. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 4770–4774.
4. Jurka, J. and Smith, T. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 4775–4778.
5. Quentin, Y. (1988) *J. Mol Evol.* **27**, 194–202.
6. Labuda, D. and Striker, G. (1989) *Nucleic Acids Res.* **17**, 2477–2491.
7. Matera, A.G., Hellmann, U. and Schmid, C.W. (1990) *Mol. Cell Biol.* **10**, 5424–5432.
8. Matera, A.G., Hellmann, U., Hintz, M.F. and Schmid, C.W. (1990) *Nucleic Acids Res.* **18**, 6019–6023.
9. Batzer, M.A. and Deininger, P.L. (1991) *Genomics*, **9**, 481–487.
10. Batzer, M.A., Kilroy, G.E., Richard, P.E., Shaikh, T.H., Desselle, T.D., Hoppens, C.L. and Deininger, **P.L. (1990)** *Nucleic Acids Res.* **18**, 6793–6798.
11. Elder, R.T., Pan, J., Duncan, C.H. and Weissman, S.M. (1981) *Nucleic Acids Res.* **9**, 1171–1189.
12. Perez-Stable, C., Ayres, T.M. and Shen, C.K.J. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 5291–5295.
13. Paulson, E.K. and Schmid, C.W. (1986) *Nucleic Acids Res.* **14**, 6145–6158.
14. Tomilin, N.V. and Bozhkov, V.M. (1989) *FEBS Letters*, **251**, 79–83.
15. Tomilin, N.V., Iguchi-Ariga, S.M.M. and Ariga, H. (1990) *FEBS Letters*, **263**, 69–72.
16. Lobo, S.M. and Hernandez, N. (1989) *Cell*, **58**, 55–67.
17. Lobo, S.M., Ifill, S. and Hernandez, N. (1990) *Nucleic Acids Res.* **18**, 2891–2899.
18. Garner, M.M. and Revzin, A. (1981) *Nucleic Acids Res.* **9**, 3047–3060.
19. Fried, M.and Crothers, D.M. (1981) *Nucleic Acids Res.* **9**, 6505–6525.
20. Singh, J. and Dixon, G.H. (1990) *Biochemistry*, **29**, 6295–6302.
21. Schreiber, E., Matthias, P., Muller, M.M. and Schaffner, W. (1989) *EMBO J.* **7**, 4221–4229.
22. Schmid, C.W. (1991) *Nucleic Acids Res.*, 19, 5613–5617.
23. Prendergast, G.C. and Ziff, E.B. (1991) *Science*, **251**, 186–189.
24. Zhang, X.-Y., Asiedu, C.K., Supakar, P.C., Khan, R., Ehrlich, K. C. and Ehrlich, M. (1990) *Nucleic Acids Res.* **18**, 6253–6260.
25. Watt, F. and Molloy, P.L. (1988) *Nucleic Acids Res.* **16**, 1471–1486.
26. Tremethick, D., and Molloy, P.L. (1986) *J. Biol. Chem.* **261**, 6986–6992.
27. Holmquist, G. (1989) *J. Mol. Evol* . **28**, 469.
28. Chesnokov, I., Bozhkov, V. Popov, B. and Tomilin, N. (1991) *Biochem. Biophys. Res. Commun.* **178** 613–61
29. Smith, S. and Stillman, B. (1991) *J. Biol. Chem.* **266**, 12041–12047.