

Multivalent DNA-binding properties of the HMG-I proteins

(HMG-Y/HMGI-C/AT-hook/interferon- β /chromatin)

JOSEPH F. MAHER* AND DANIEL NATHANS†

Departments of Molecular Biology and Genetics, and Medicine, Howard Hughes Medical Institute, Johns Hopkins University School of Medicine, 725 North Wolfe Street, Baltimore, MD 21205

Contributed by Daniel Nathans, March 13, 1996

ABSTRACT HMG-I proteins are DNA-binding proteins thought to affect the formation and function of transcription complexes. Each protein contains three DNA-binding motifs, known as AT-hooks, that bind in the minor groove of AT tracts in DNA. Multiple AT-hooks within a polypeptide chain should contact multiple AT tracts, but the rules governing these interactions have not been defined. In this study, we demonstrate that high-affinity binding uses two or three appropriately spaced AT tracts as a single multivalent binding site. These principles have implications for binding to regulatory elements such as the interferon β enhancer, TATA boxes, and serum response elements.

The HMG-I family of high mobility group chromosomal proteins consists of three members: HMG-I and HMG-Y, which are alternatively spliced products of a single gene (1–3), and HMGI-C, which is the product of a separate gene (4, 5). There is evidence that HMG-I(Y) is involved in the transcriptional regulation of genes, the best characterized example being the interferon β gene (6–17). Specific biochemical functions of HMGI-C have not been elucidated, but recent reports of an HMGI-C knockout in the mouse causing the *pygmy* phenotype (18) and the implication of HMGI-C gene rearrangements in lipomas and other benign mesenchymal tumors (19) suggest an important role for this protein in cell growth and differentiation. Consistent with such a role is the activation of HMG-I genes as part of the delayed early response to growth factors (20).

The HMG-I proteins are DNA-binding proteins, and the DNA-binding motif is known as an AT-hook (21), which has been shown to bind in the minor groove of AT tracts of double-stranded DNA (22, 23). Footprinting studies have shown that many AT tracts of 4, 5, 6, or more bp can be bound by HMG-I proteins (22). However, of the multitude of AT tracts in cellular DNA, such as TATA boxes, homeodomain-binding sites, serum response elements, matrix attachment regions, and many others, it is not clear which among them are high-affinity HMG-I protein-binding sites of functional significance.

Since each HMG-I protein molecule contains three AT-hooks, each one is presumably able to mediate contact with an AT tract in DNA. This suggests the likelihood that high-affinity binding would involve multivalent binding of a single polypeptide to multiple AT tracts in close proximity to each other, although this form of binding has not been well defined. The purpose of the studies described in this report was to define multivalent, high-affinity binding sites for the HMG-I proteins.

MATERIALS AND METHODS

Constructs. HMG-I, HMG-Y, and HMGI-C were cloned by PCR in pBX/*Bsp*HI. The resulting clones were verified by

sequencing. HMGI-C was then subcloned into pET-3d (Novagen) and HMG-I and -Y were subcloned into pET-28d (Novagen), using *Escherichia coli* strain DH5 α (GIBCO/BRL) to produce the plasmids pET/HMG-C, pET/HMG-I, and pET/HMG-Y, respectively.

Protein Purification. The above plasmids were transferred into *E. coli* strain BL-21(DE3) (Novagen) and protein expression was induced with isopropyl thiogalactoside at 1 mM for 3 hr. Cells were harvested by centrifugation, washed in ice-cold buffer [50 mM Tris, pH 7.5/10 mM magnesium acetate/1 mM EDTA/10% glycerol/1 mM DTT/23 μ g/ml phenylmethylsulfonyl fluoride (PMSF)], and stored overnight at -80°C . The pellet was resuspended in ice-cold lysis buffer [50 mM Tris, pH 7.4/500 mM NaCl/20 mM MgCl₂/1 mM DTT/23 μ g/ml PMSF], lysozyme was added at 0.1 mg/ml, and the cells were sonicated on ice. After centrifugation at $400 \times g$, one-half volume of acetone at room temperature was added to the supernatant, and the mixture was immediately centrifuged at $4000 \times g$ for 10 min at 4°C . To the supernatant, another one-half volume of ice-cold acetone was added, and after 20 min on ice, the mixture was centrifuged at $4000 \times g$. The air-dried pellet was then resuspended in HME buffer [20 mM Hepes, pH 7.8/5 mM MgCl₂/0.2 mM EDTA/23 μ g/ml PMSF] and applied to an SP-Sephadex column with a gravity flow rate. The column was eluted at 4°C with a step gradient of KCl in HME buffer. For HMG-I and -Y, fractions containing electrophoretically identified HMG protein were pooled and applied to a Phosphocellulose column at 4°C and eluted with a step gradient of KCl in HE buffer [20 mM Hepes, pH 7.8/0.2 mM EDTA]. Fractions containing HMG proteins were concentrated and buffer-exchanged into HK buffer [20 mM Hepes, pH 7.8/20 mM KCl/0.2 mM EDTA/0.1% Nonidet P-40], and protein concentrations were measured with the Bradford assay (Bio-Rad). Purified, bacterially expressed NF κ B p50 subunit was a gift of C. Rosen and C. Kunsch (Human Genome Science).

Oligonucleotide Synthesis and Labeling. Oligonucleotides were chemically synthesized using standard methodology. Oligonucleotides used in the experiments (see Figs. 1, 2, and 3) were double-stranded 60-mers and contained homopolymeric AT tracts separated and flanked by GC-rich stretches. For example, the oligonucleotide designated T5–6–T5 had the following sequence in one strand: GGACTCCAGGTCCAG-GACCGCGTTTTTCGCGCGTTTTTCGCGGGAGGT-CCAGCTGTCCACCTCC.

Other oligonucleotides were of the same general type. Most oligonucleotides had a common 3' sequence that is complementary to the primer 5'-GGAGGTGGAC-3', which was used

Abbreviations: HMG, high mobility group; EMSA, electrophoretic mobility shift assay; PRD, positive regulatory domain; NRD, negative regulatory domain.

*Present address: Department of Medicine, University of Mississippi School of Medicine and Veterans Affairs Medical Center—Research Service (151), Room R-417A, 1500 East Woodrow Wilson Drive, Jackson, MS 39216.

†To whom reprint requests should be addressed.

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

to prime synthesis of double-stranded DNA with the Klenow fragment of DNA polymerase I. Each oligonucleotide was then purified on a 12% polyacrylamide gel in 1× TBE buffer (89 mM Tris/89 mM borate/2 mM EDTA). In the case of the positive regulatory domain II (PRD II)/negative regulatory domain I (NRD I) oligonucleotides, both strands were chemically synthesized and annealed by slow cooling of an equimolar mixture from 100°C, then purified on a 12% polyacrylamide gel as above (for their sequences, see Fig. 4). Double-stranded oligonucleotides were end-labeled with polynucleotide kinase and [γ -³²P]ATP at 6000 Ci/mmol (1 Ci = 37 GBq). The resultant reaction mixtures were heat-inactivated and purified over NICK columns (Pharmacia). In all cases, the oligonucleotides to be compared had specific activities within 12% of each other.

Electrophoretic Mobility Shift Assay (EMSA). Unless otherwise noted, binding reactions were carried out at room temperature (22°C) in a mixture containing 10 mM Tris (pH 7.5), 50 mM KCl, 0.5 mM MgCl₂, 0.1 mM EDTA, 250 μg/ml acetylated BSA, and 5% glycerol in reaction volumes of 30 ml. After incubation for 30 min, the binding mixtures were loaded onto pre-electrophoresed 6% polyacrylamide gels in 0.4 × TBE, run at 4°C and 300 V, and autoradiographed.

RESULTS

Effect of Spacing Between Two AT Tracts on Binding of HMG-I, HMG-Y, and HMGI-C. To look for multivalent binding, we used EMSAs with purified, recombinant HMG-I, HMG-Y, and HMGI-C and synthetic oligonucleotides with two AT tracts in which the spacing between the tracts was systematically varied. Direct binding to labeled oligonucleotides rather than band competition was used to assess the stoichiometry of the bound complexes. Fig. 1 shows the pattern of binding to oligonucleotides containing two homopolymeric T5·A5 tracts with spacing between them of 0, 2, 4, 6, 8, 10, 12, 14, 16, 20, and 30 bp; to an oligonucleotide containing a single T5·A5 tract; and to an oligonucleotide containing no AT tract. As can be seen in Fig. 1, all HMG-I proteins bind with higher affinity to two AT tracts with a spacing of up to 8 or 10 bp, with a peak of affinity occurring with spacing of 6–8 bp. The pattern of high-affinity binding is somewhat more restricted with HMG-I than with HMG-Y and HMGI-C.

Effect of AT Tract Size on Binding. Fig. 2 demonstrates that the affinity of binding also increases with the size of the AT tract when a single T8·A8 tract is compared with a T5·A5 tract or when two spaced tracts are compared. For each protein, binding to two T5·A5 tracts with a spacing of 6 bp is of higher affinity than binding to a single T8·A8 tract (or the single T10·A10 tract in T5–0-T5, as seen in Fig. 1). However, the pattern of binding to AT tracts with different spacing between them is essentially the same whether T8·A8 or T5·A5 tracts are used (data not shown).

The results described above do not establish whether AT hooks of a single HMG monomer bind to the two AT tracts of an oligonucleotide or whether HMG dimers are the binding species. To distinguish between these two possibilities, we used an equimolar mixture of HMGI-C and GST/HMGI-C (a glutathione S-transferase–HMG fusion protein) to look for heteromeric protein-oligonucleotide complexes with mobilities between those of the HMGI-C and GST/HMGI-C complexes. Since no such heteromeric complexes were detected (data not shown), we infer that each complex is likely to be monomeric.

Two or Three Optimally Spaced AT Tracts Form a Single, Multivalent Binding Site. We next determined more directly whether two optimally spaced AT tracts form a single binding site for HMG-I proteins. Fig. 3A and B demonstrates that two optimally spaced AT tracts behave as a single, high-affinity binding site, whereas with greater spacing, they behave as two

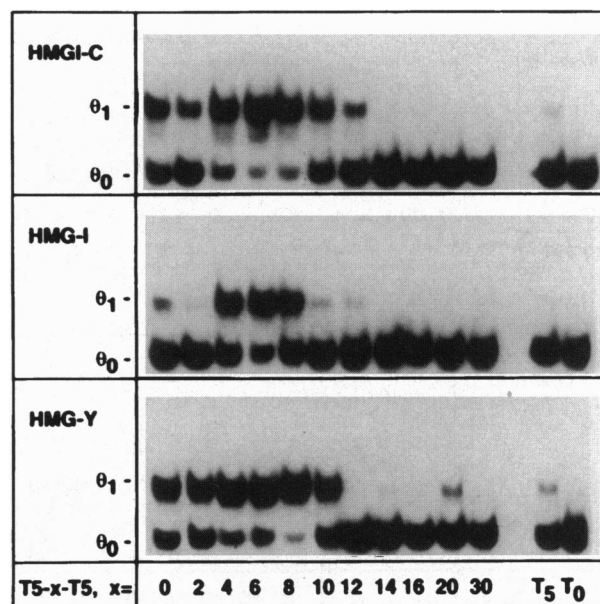


Fig. 1. EMSA of purified recombinant HMG-I, HMG-Y, and HMGI-C binding to double-stranded oligonucleotides containing two homopolymeric T5·A5 tracts separated by variable spacing. The spacing in base pairs between the two T5·A5 tracts is noted at the bottom. Binding to oligonucleotide T₅ (with a single T5·A5 tract) and oligonucleotide T₀ (with no AT tract) are present for comparison. θ₀ designates the unbound oligonucleotide, and θ₁ designates a singly liganded protein-oligonucleotide complex. The concentrations of HMG-I, HMG-Y, and HMGI-C are 1000, 250, and 100 pM, respectively, and the concentration of each oligonucleotide was 33 pM. Binding reactions were done at 22°C for 30 min. The faint band seen below the main complex in experiments with HMGI-C is probably due to the binding of a proteolytic product formed during protein purification, as seen by Coomassie staining or specific immunostaining of the purified HMGI-C preparation following gel electrophoresis.

low-affinity binding sites. In Fig. 3A, saturation curves demonstrate that the T5–6-T5 oligonucleotide saturates as a singly liganded site (θ₁ complex), whereas the T5–12-T5 and T5–20-T5 oligonucleotides readily form a doubly liganded complex (θ₂ complex). Fig. 3B demonstrates this phenomenon with a broader range of spacings between two T8·A8 tracts at a saturating concentration of HMGI-C. As can be seen, high-affinity sites with spacing of ≤8 bp are resistant to double saturation, whereas more widely spaced, low-affinity AT tracts show much more double saturation. (The relatively low amount of θ₂ complex seen with high-affinity sites is similar to

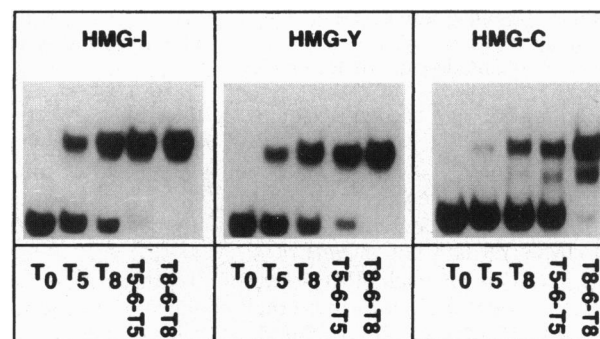


Fig. 2. EMSA of HMG-I, HMG-Y, and HMGI-C binding to oligonucleotides containing a single T5·A5 or T8·A8 tract, or with two T5·A5 or T8·A8 tracts separated by 6 bp. Oligonucleotide T₀ with no AT tract was included for comparison. The concentrations of proteins used was 500, 200, and 20 pM for HMG-I, HMG-Y, and HMGI-C, respectively, and oligonucleotides were each 14 pM. Binding reactions were done at 22°C for 30 min.

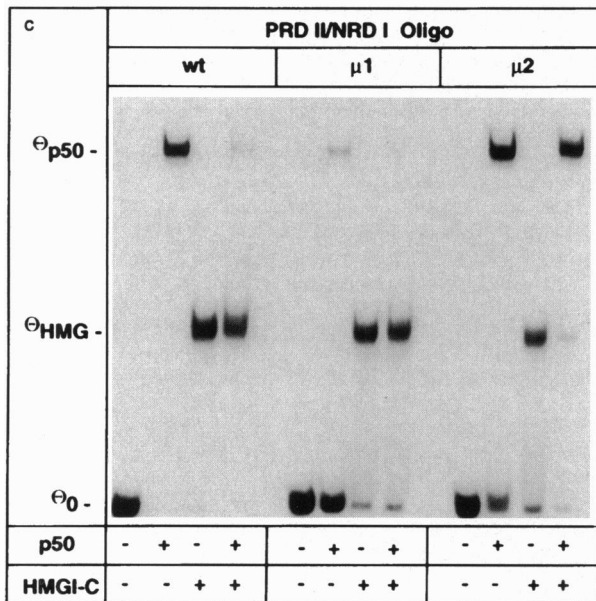
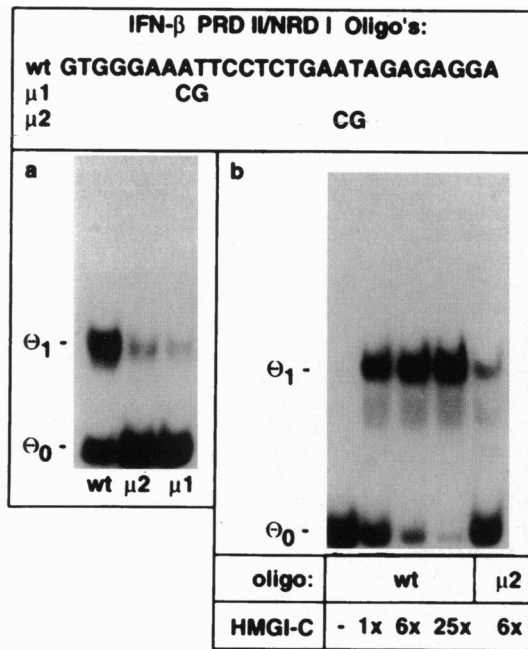


FIG. 4. (A) EMSA analysis of the binding of HMGI-C to the interferon β PRD II/NRD I wild-type element and the μ 1 and μ 2 mutant elements containing the double base pair substitutions noted. The concentration of HMGI-C was 2.5 nM and that of the oligonucleotide was 200 pM. (B) EMSA analysis of binding of increasing concentrations of HMGI-C to the PRD II/NRD I wild-type site and to the μ 2 mutant site for comparison. A 1 \times concentration of HMGI-C was 0.8 nM, and the oligonucleotide was at 200 pM. (C) EMSA analysis of the binding of saturating concentrations of HMGI-C and p50 to the PRD II/NRD I wild-type and μ 1 and μ 2 mutant sites. θ_{HMG} designates the HMGI-C/DNA complex and θ_{p50} designates the p50/DNA complex. HMGI-C and labeled oligonucleotide were preincubated together on ice for 10 min before the addition of p50. In A, B, and C, binding reactions were carried out at 0°C for 30 min.

loss of multivalent high-affinity binding when the spacing between the two AT tracts is greater than 10 bp, we suppose that this spacing is beyond the maximal span of the AT-hooks in the polypeptide chain. The more restricted pattern of multivalent binding by HMGI-I seen in Fig. 1 could be explained by its unique amino acid sequence, which either alters flexibility of the polypeptide chain between AT-hooks or

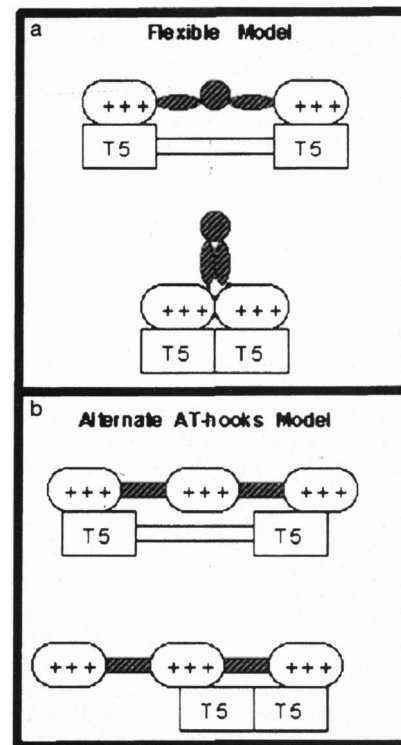


FIG. 5. Schematic models of multivalent binding properties of HMGI-I proteins. (A) Flexible model. (B) Alternate AT-hooks model. The ovals containing the +++ symbols represent the positively charged AT-hooks. The boxes containing the T5 symbol represent AT tracts in double-stranded DNA.

diminishes the DNA-binding activity of one of the alternate AT-hooks. Further analysis using altered HMGI-I, HMGI-Y, and HMGI-C proteins generated by site-directed mutagenesis and structural studies of protein-DNA complexes will be needed to further test these models.

Multivalent binding of HMGI-I proteins as explored in this paper has important implications for the binding of HMGI-I proteins within regulatory regions of DNA. It predicts that single AT tracts, such as in TATA boxes, serum response elements, and homeodomain-binding sites, would by themselves be univalent, low-affinity binding sites for HMG proteins. However, appropriate spacing of a second (or third) AT tract could convert a site into a multivalent, high-affinity HMG binding site. Based on the results shown in Fig. 4, we speculate that one effect of HMGI-I(Y) binding to the interferon β promoter could be to displace a putative transcriptional repressor molecule/complex bound at NRD I, thereby relieving the preinduction repression of the gene following virus infection (25). NRD I actually overlaps PRD II and includes both AT tracts (24). Such an effect of HMGI-I(Y) could help accentuate the off/on switch in the transcriptional activation of this gene. It is interesting to note the published data showing that although non-HMG binding NF κ B sites, such as the Ig- κ B element, when multimerized in front of a minimal promoter, confers virus inducibility of a reporter gene, this same element cannot function in virus induction when substituted for the PRD II NF κ B site in the context of the interferon β enhancer also containing NRD I (6). This substitution would make the HMGI-I(Y) binding site in NRD I a univalent, low-affinity site.

In addition to the finding of multivalent binding of HMGI-I proteins in the interferon β enhancer, we have preliminary data showing that the HMGI-I proteins bind with high affinity to the downstream JunB enhancer, which has a serum response element and a second AT tract spaced 5 bp away from the AT

tract in the serum response element (26), whereas they bind with low affinity to mutant enhancers in which either AT tract has been disrupted. In contrast, HMG-I proteins bind with low affinity to the c-Fos serum response element, which does not have an appropriately spaced second AT tract (27). Multivalent binding of HMG-I proteins could also have implications for DNA structural changes, such as DNA bending or bend reversal (28), which can affect the binding of other proteins to the same DNA element. Finally, it is worth noting that in one of the original descriptions of HMG-I, in which HMG-I was defined as a protein that bound to a satellite DNA, it was pointed out that three binding sites in this repetitive DNA sequence could come into proximity when wrapped around a nucleosome (29). If the multivalent binding properties of the HMG-I proteins reported here are extended to the three-dimensional space of the nucleosome, then three suitably spaced AT tracts in DNA that is wrapped around a nucleosome might be bound by the three AT-hooks of a single HMG-I protein.

1. Lund, T., Holtlund, J., Fredriksen, M. & Laland, S. G. (1983) *FEBS Lett.* **152**, 163–167.
2. Johnson, K. R., Lehn, D. A. & Reeves, R. (1989) *Mol. Cell. Biol.* **9**, 2114–2123.
3. Friedmann, M., Holth, L. T., Zoghbi, H. & Reeves, R. (1993) *Nucleic Acids Res.* **21**, 4259–4267.
4. Manfioletti, G., Giancotti, V., Bandiera, A., Buratti, E., Sautiere, P., Cary, P., Crane-Robinson, C., Coles, B. & Goodwin, G. H. (1991) *Nucleic Acids Res.* **19**, 6793–6797.
5. Patel, U. A., Bandieri, A., Manfioletti, G., Giancotti, V., Chau, K.-Y. & Crane-Robinson, C. (1994) *Biochem. Biophys. Res. Commun.* **201**, 63–70.
6. Thanos, D. & Maniatis, T. (1992) *Cell* **71**, 777–789.
7. Du, W., Thanos, D. & Maniatis, T. (1993) *Cell* **74**, 887–898.
8. Du, W. & Maniatis, T. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 11318–11322.
9. Thanos, D. & Maniatis, T. (1995) *Cell* **83**, 1091–1100.
10. John, S., Reeves, R. B., Lin, J.-X., Child, R., Leider, J. M., Thompson, C. B. & Leonard, W. J. (1995) *Mol. Cell. Biol.* **15**, 1786–1796.
11. Neish, A. S., Read, M. A., Thanos, D., Pine, R., Maniatis, T. & Collins, T. (1995) *Mol. Cell. Biol.* **15**, 2558–2569.
12. Whitley, M. Z., Thanos, D., Read, M. A., Maniatis, T. & Collins, T. (1994) *Mol. Cell. Biol.* **14**, 6464–6475.
13. Leger, H., Sock, E., Renner, K., Grummt, F. & Wegner, M. (1995) *Mol. Cell. Biol.* **15**, 3738–3747.
14. Fashena, S. J., Reeves, R. & Ruddle, N. H. (1992) *Mol. Cell. Biol.* **12**, 894–903.
15. Chuvpilo, S., Schomberg, C., Gerwig, R., Heinfling, A., Reeves, R., Grummt, F. & Serfling, E. (1993) *Nucleic Acids Res.* **21**, 5694–5704.
16. Lewis, H., Kaszubska, W., DeLamarter, J. F. & Whelan, J. (1994) *Mol. Cell. Biol.* **14**, 5701–5709.
17. Abdulkadir, S. A., Krishna, S., Thanos, D., Maniatis, T., Strominger, J. L. & Ono, S. J. (1995) *J. Exp. Med.* **182**, 487–500.
18. Zhou, X., Benson, K. F., Ashar, H. R. & Chada, K. (1995) *Nature (London)* **376**, 771–774.
19. Ashar, H. R., Schoenberg Fezjo, M., Tkachenko, A., Zhou, X., Fletcher, J. A., Weremowicz, S., Morton, C. C. & Chada, K. (1995) *Cell* **82**, 57–65.
20. Lanahan, A., Williams, J. B., Sanders, L. K. & Nathans, D. (1992) *Mol. Cell. Biol.* **12**, 3919–3929.
21. Reeves, R. & Nissen, M. S. (1990) *J. Biol. Chem.* **265**, 8573–8582.
22. Solomon, M. J., Strauss, F. & Varshavsky, A. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 1276–1280.
23. Geierstanger, B. H., Volkman, B. F., Kremer, W. & Wemmer, D. E. (1994) *Biochemistry* **33**, 5347–5355.
24. Goodbourn, S. & Maniatis, T. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 1447–1451.
25. Goodbourn, S., Burstein, H. & Maniatis, T. (1986) *Cell* **45**, 601–610.
26. Perez-Albuerne, E. D., Schatteman, G., Sanders, L. K. & Nathans, D. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 11960–11964.
27. Treisman, R. (1992) *Trends Biochem. Sci.* **17**, 423–426.
28. Falvo, J. V., Thanos, D. & Maniatis, T. (1995) *Cell* **83**, 1101–1111.
29. Strauss, F. & Varshavsky, A. (1984) *Cell* **37**, 889–901.