Cell, Vol. 37, 889–901, July 1984, Copyright © 1984 by MIT 0092-8674/84/070889-12 $02.00/0

A Protein Binds to a Satellite DNA Repeat at Three Specific Sites That Would Be Brought into Mutual Proximity by DNA Folding in the Nucleosome

François Strauss and Alexander Varshavsky
Department of Biology
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

Summary

Using a generally applicable assay for specific DNA-binding proteins in crude extracts, we have detected and purified an HMG-like nuclear protein from African green monkey cells that preferentially binds to the 172 bp repeat of α-satellite DNA (α-DNA). DNAase I footprinting with the purified protein detects three specific binding sites (I–III) per α-DNA repeat. Site II is 145 bp (one core nucleosome length) from site III on the adjacent α-DNA repeat, while site I lies midway between sites II and III. In the α-nucleosome phasing frame corresponding with this arrangement, sites I–III would be brought into mutual proximity by DNA folding in the nucleosome. This phasing frame is identical with the preferred frame detected previously in isolated chromatin. Our results suggest that this new and abundant protein recognizes a family of short, related nucleotide sequences found not only in α-DNA but also throughout the genome, and that functions of this protein are mediated through its nucleosome-positioning activity. Such nucleosome-positioning proteins may underlie the sequence specificity of both nucleosome arrangements and higher order chromatin structures.

Introduction

Eucaryotic chromosomes contain nucleotide sequences of lengths from about 10 to more than 10^3 base pairs (bp) that are repeated thousands to millions of times per haploid genome. Highly repeated DNA is arranged largely in long tandem arrays ("satellite" DNA) and is found mostly within centromeric and telomeric chromosomal regions, where it comprises the bulk of constitutive heterochromatin (reviewed by Peacock et al., 1978; John and Miklos, 1979; Brutlag, 1980; Hilliker et al., 1980; Singer, 1982; see also Gatty et al., 1983).

No function of highly repeated, tandemly arranged DNA sequences has been identified. Examples of the near total elimination of the satellite DNA from precursors of somatic (but not of germ-line) nuclei during development in a variety of species (Beerman and Meyer, 1980; Klobutcher et al., 1981; Cavalier-Smith, 1982; Streeck et al., 1983; Müller et al., 1983) suggest that any essential function of the bulk of satellite DNA may be confined to germ cell lineages.

Although it has been suspected for many years that distinctive properties of heterochromatin, such as its transcriptional inactivity and highly compact structure, are due in part to the presence of heterochromatin-specific proteins, no direct evidence has been available until recently. The first apparently heterochromatin-specific nucleosomal protein identified was D1, an abundant Drosophila nuclear protein isolated by Rodriguez-Alfageme et al. (1980) and shown by Levinger and Varshavsky (1982a, 1982b) to be a stoichiometric component of Drosophila melanogaster nucleosomes containing (A+T)-rich satellite DNA. ProteinD1 preferentially binds to (A+T) tracts of double-stranded DNA in vitro, and is present in (A+T)-rich satellite nucleosomes in addition to core histones (Levinger and Varshavsky, 1982a, 1982b; Varshavsky et al., 1983). Potentially analogous DNA-binding proteins, but with undefined nucleosome-binding properties, have been detected in other eucaryotic cells as well (Hsieh and Brutlag, 1978; Bennet et al., 1982; Avilla et al., 1983; Garreau and Williams, 1983).

To see whether there exists a family of sequence-specific nucleosomal proteins, we have searched for a protein specific for another tandemly repetitive DNA, the α-satellite DNA (α-DNA) of the African green monkey; the 172 bp repeat of α-DNA comprises 15% to 20% of the green monkey genome (Rosenberg et al., 1978). By two-dimensional hybridization mapping of nucleosomes (Levinger et al., 1981), we have shown that the nucleosomes containing α-DNA (α-nucleosomes) behave as if specifically associated with an additional protein (Wu et al., 1983), in striking similarity to the previously obtained, analogous data for D1 protein in Drosophila (Levinger and Varshavsky, 1982a, 1982b). To isolate this protein from green monkey CV-1 cells, we devised a generally applicable assay (called a "band-competition" assay) that allows the detection of specific DNA-binding proteins in a crude extract and the monitoring of their subsequent purification. Using this assay with CV-1 cells, we have detected and purified an abundant HMG-like protein ("α-protein") that binds to α-DNA in vitro at three specific sites per α-DNA repeat. The striking arrangement of these sites within the α-DNA repeat suggests a specific α-nucleosome phasing frame in which bending of DNA in the phased α-nucleosome would bring all three sites into mutual proximity on one side of the nucleosome, close to the DNA "exit" point.

Our results suggest the existence of a new class of DNA-binding proteins that recognize families of short, related nucleotide sequences throughout the genome and function as nucleosome-positioning proteins. We present simple quantitative arguments that even a moderately sequence-specific protein can be preferentially bound in vivo to its cognate DNA sites if the latter are abundant enough. This suggests that not only α-protein but also other abundant chromosomal proteins may function via sequence-specific DNA recognition. Nucleosome-positioning, or phasing, proteins may underlie the sequence specificity of both nucleosome arrangements and higher order chromatin structures. We suggest a temporally controlled, phasing-protein-mediated positioning of nucleosomes at or near DNA sites required for initiation of transcription as a new genetic regulatory mechanism and briefly discuss two specific systems, mammalian X-chromosome inacti-

vation and regulation of 5S RNA genes in amphibian development, as potential examples of such a mechanism.

Results

Detection of α-DNA-Binding Protein in a Crude Nuclear Extract by the Band-Competition Assay

Purified nuclei from green monkey CV-1 cells were extracted with 0.35 M NaCl (see Experimental Procedures). An aliquot of the extract was added to the ^{32}P-labeled, 172 bp α-DNA monomer, and the sample was electrophoresed on a low-ionic-strength 4% polyacrylamide gel (Figure 1A, lane b). In the absence of added 0.35 M NaCl extract the α-DNA monomer migrated in the gel as a discrete band (lane a), while in the presence of the extract most of the DNA failed to enter the gel (lane b), the expected result of adding a vast excess of DNA-binding proteins to a small amount of naked DNA. Thus, when a sufficiently large excess of unlabeled E. coli competitor DNA was added to the extract together with the ^{32}P-labeled α-DNA monomer, the latter migrated with the mobility of naked α-DNA (lanes k–r). At intermediate ratios of E. coli DNA to α-DNA several discrete α-DNA–protein complexes were observed (lanes f–j); the most rapidly migrating α-DNA–protein complex (complex I) was still seen at an E. coli DNA/α-DNA weight ratio of ~270 (lane j), suggesting that a protein of complex I had a considerably higher affinity for α-DNA than for E. coli DNA.

In a different test for the presence of an α-DNA-binding protein in the crude extract, a series of band-competition assays was carried out with a randomly selected, 180 bp ^{32}P-DNA fragment (derived from pBR322 DNA by digestion with Msp I) instead of the 172 bp α-DNA monomer (see Experimental Procedures and Figure 1B). Although the 180 bp DNA fragment of pBR322 did form complexes with proteins in the CV-1 nuclear extract (Figure 1B), they were distinct in two specific respects from analogous complexes formed by the 172 bp α-DNA monomer. First, while α-DNA–protein complexes were detectable up to an E. coli DNA/α-DNA weight ratio of ~270 (Figure 1A, lane j), analogous complexes containing the 180 bp pBR322 DNA fragment disappeared at a reproducibly lower E. coli DNA/pBR322 DNA weight ratio of ~70 (Figure 1B, lane h). Second, while α-DNA–protein complexes, such as complexes I and II (Figure 1A, lanes h–j; see also Figure 2C), were sufficiently stable to form discrete electrophoretic bands, analogous DNA–protein complexes containing the 180 bp pBR322 DNA fragment (Figure 1B, lanes i–h) reproducibly formed smeared bands, suggesting less stable complexes.

Taken together, the above results strongly suggest that the CV-1 nuclear extract contains a protein or proteins that preferentially bind to the α-DNA repeat ("α-protein").

Use of the Band-Competition Assay to Monitor Purification of α-Protein

In the first purification step, the 0.35 M NaCl extract of CV-1 nuclei was fractionated by phosphocellulose chromatography; aliquots from even-numbered fractions were used to detect α-DNA-binding activity (Figures 2A and 2B) and to visualize the distribution of proteins by polyacrylamide–SDS gel electrophoresis and silver staining (Figures 3A–3C). In the presence of an optimal amount of E. coli competitor DNA the band-competition assay allowed the detection of a single peak of α-DNA-binding activity that eluted at ~0.5 M K-phosphate (fractions 50–54 in Figures 2A and 2B). The most abundant protein species in these fractions (α-protein) almost comigrates with the high mobility group (HMG) protein 17 in the SDS gel (Figure 3C); however, it is distinct from HMG17 (see below) and elutes

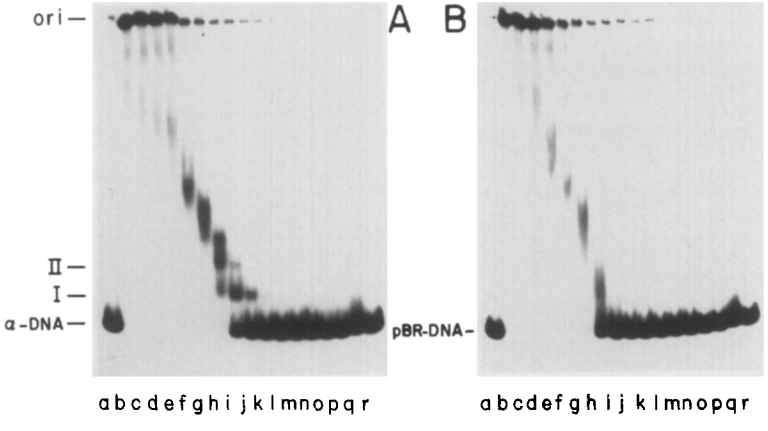


Figure 1. Detection of α-Satellite DNA-Binding Protein in a Crude Extract by the Band-Competition Assay

(A) The ^{32}P-labeled, 172 bp α-DNA monomer was mixed with a fixed amount of the 0.35 M NaCl extract of CV-1 nuclei and with increasing amounts of unlabeled sonicated E. coli competitor DNA before electrophoresis in a low-ionic-strength 4% polyacrylamide gel (see Experimental Procedures). (Lane a) α-DNA monomer (25,000 cpm, 2.5 ng of DNA) in the absence of added extract; (lanes b–r) same but in the presence of the extract (5 μl, containing ~1 μg of total protein) together with 0, 5, 11, 22, 43, 87, 173, 340, 680, 1.4×10^3, 2.8×10^3, 5.5×10^3, 1.1×10^4, 2.2×10^4, 4.4×10^4, 8.8×10^4 and 17.6×10^4 ng of E. coli DNA per assay, respectively.

(B) Same as in A except that a ^{32}P-labeled, 180 bp long DNA fragment excised from pBR322 with Msp I (positions 1484 to 1664; see Maniatis et al., 1982) was used instead of the α-DNA monomer. I and II in A indicate discrete α-DNA–protein complexes observed in the presence of an optimal excess of E. coli competitor DNA (the numbers of α-protein molecules per α-DNA fragment in I and II have not been determined); ori, electrophoretic origin.

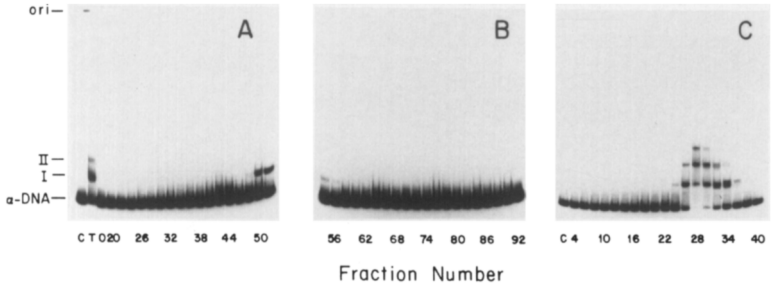


Figure 2. Monitoring α-Protein Purification by the Band-Competition Assay

(A,B) Aliquots (1 μl) of even-numbered fractions (20–92) from the phosphocellulose column (see Experimental Procedures) were tested for the presence of α-DNA-binding activity by the band-competition assay (see Experimental Procedures and the legend to Figure 1) carried out in the presence of a fixed amount of E. coli competitor DNA (250 ng of E. coli DNA per ~0.25 ng of the ^{32}P-labeled α-DNA monomer per assay).

(C) Phosphocellulose fractions (A and B) containing the α-DNA-binding activity (fractions 50–54, eluted at ~0.5 M K-phosphate) were pooled and chromatographed on a DNA–Sephacryl column (see Experimental Procedures). Fractions were assayed for the α-DNA-binding activity as described above; active fractions (27–35) eluted at ~0.3 M NaCl, and were subjected to hydroxyapatite chromatography, the last purification step (see Experimental Procedures and Figure 3E). The corresponding silver-stained protein patterns are shown in Figures 3A–3E. Designations: C, no chromatographic fractions added (control); T, same but with the total (unfractionated) extract added; O, same but with the flowthrough fraction from the phosphocellulose column added. Other designations are as in Figure 1.

from phosphocellulose after both HMG14 and HMG17 (Figure 3C). Phosphocellulose fractions 50–54 were pooled and subjected to chromatography on an E. coli DNA–Sephacryl column (Figures 2C and 3D). Figure 2C shows that the α-DNA-binding activity continued to copurify with the α-protein. In the third and last chromatographic step, the pooled α-protein fractions 27–35 from the DNA–Sephacryl column (Figure 2C) were chromatographed on hydroxyapatite to yield an apparently homogeneous α-protein (Figure 3E, fraction 17). The purified α-protein preferentially bound to α-DNA in a band-competition assay (Figure 4) analogous to the one carried out with the crude nuclear extract (Figure 1), and was apparently the only abundant nuclear protein in cultured green monkey cells that possessed this property (see Figures 2 and 3).

α-Protein Is an HMG-like Nuclear Protein Comparable in Abundance to HMG17 and HMG14

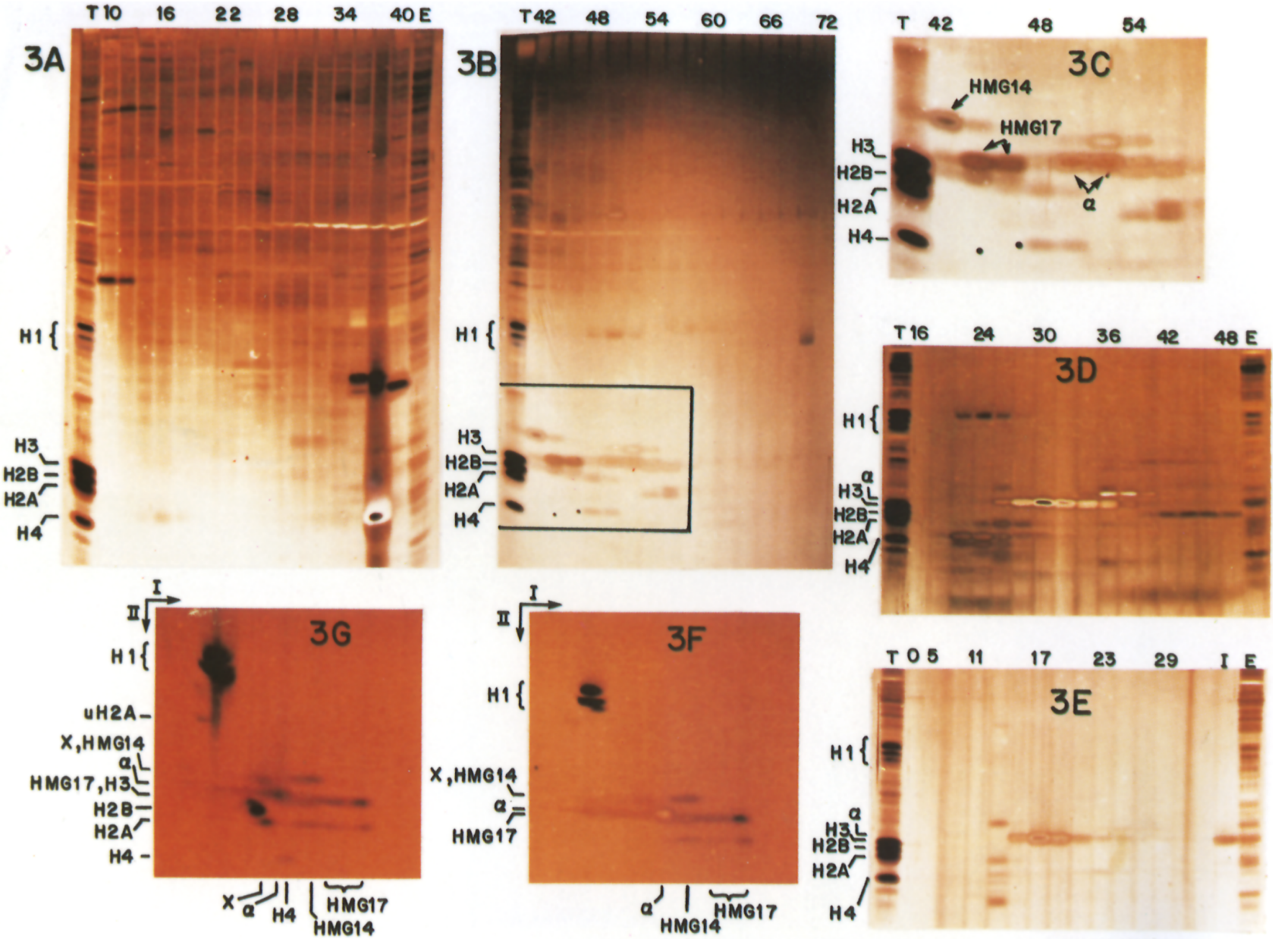
To characterize α-protein and its relationship to HMG proteins further (see Figure 3C), the 0.35 M NaCl extract of CV-1 nuclei was treated with 5% (w/v) CCl_3COOH; the 5% CCl_3COOH-soluble proteins, which include HMG14, HMG17, and histone H1, were resolved by two-dimensional electrophoresis (first dimension, acetic acid–urea; second dimension, SDS) and visualized by silver staining (Figure 3F). Purified α-protein, processed and fractionated identically in a separate experiment, was used as a marker (not shown). α-Protein was not only soluble in 5% CCl_3COOH as were HMG14 and HMG17; in addition, it almost comigrated with HMG17 (an 89 residue protein; Walker, 1982) in the second (SDS) electrophoretic dimension (Figures 3C, 3F, and 3G), and migrated only slightly slower than HMG14 in the first (acetic acid–urea) dimension (Figures 3F and 3G). Thus α-protein can be confused with either HMG14 or HMG17 in unidimensional electrophoresis, although it is clearly distinguishable from both HMG14 and HMG17 in the two-dimensional electrophoretic patterns (Figures 3F and 3G). This observation may in part account for the fact that α-protein was not originally identified as a distinct and abundant HMG protein (but see Discussion).

α-Protein shares the distinctive amino acid composition characteristic of HMG proteins but differs from HMG14 and HMG17 in particular in its much higher content of serine residues, precluding the possibility that α-protein is a modified form of either HMG14 or HMG17 (J. Smart, F. Strauss, and A. Varshavsky, unpublished data).

The apparent relative abundances of HMG14, HMG17, and α-protein did not differ much between the 0.35 M NaCl nuclear extract, which contained little if any of the core histones, and the 0.2 N HCl nuclear extract, which contained all of the histone species (Figures 3F and 3G and data not shown). The molar content of 0.35 M NaCl–extractable α-protein in CV-1 nuclei is between 5% and 15% of the molar content of histone H4—i.e., between 1 and 3 molecules of α-protein per 10 nucleosomes (Figure 3 and data not shown).

DNAase I Footprinting with Purified α-Protein Detects Three Specific Binding Sites on α-DNA

A double-stranded DNA probe containing a cloned α-DNA monomer was labeled with ^{32}P at one end (see Experimental Procedures). Increasing amounts of the purified α-protein were incubated with probe DNA, then subjected to partial cleavage with DNAase I. Regions protected from DNAase appeared as a gap, or footprint (Galas and Schmitz, 1978), in an otherwise uninterrupted ladder of cleavage products (Figure 5). Purified α-protein protected the α-DNA monomer at three specific sites whose positions and lengths (approximately 7 bp; see below) could be deduced from comparison with a Maxam–Gilbert G+A sequence ladder run alongside the footprint (Figure 5). Sites II and III contain inverted 7 bp repeats of each other (5′-AAATATC and 5′-GATATTT), while site I contains a weakly homologous sequence (5′-TTAATTC) (Figure 5). The 3′-terminal C residue in sites I, II, and III is strongly and


3A
T 10 16 22 28 34 40 E
H1
H3
H2B
H2A
H4
3B
T 42 48 54 60 66 72
H1
H3
H2B
H2A
H4
3C
T 42 48 54
HMG14
HMG17
H3
H2B
H2A
H4
α
3D
T 16 24 30 36 42 48 E
H1
α
H3
H2B
H2A
H4
3E
T 0 5 11 17 23 29 I E
H1
α
H3
H2B
H2A
H4
3F
I
II
H1
X, HMG14
α
HMG17
α
HMG14
HMG17
3G
I
II
H1
uH2A
X, HMG14
α
HMG17, H3
H2B
H2A
H4
X
α
H4
HMG17
HMG14

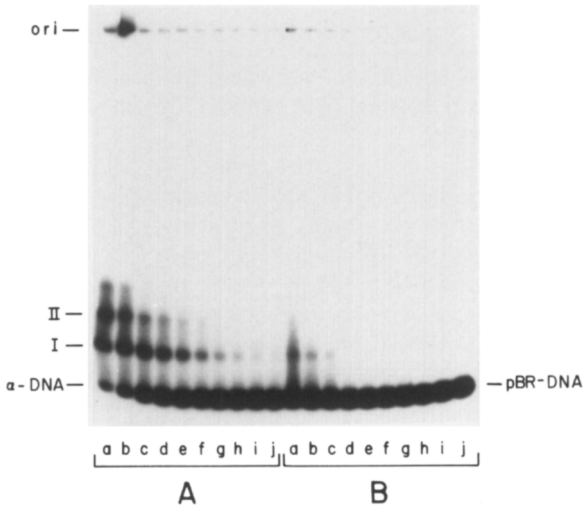


Figure 4. Binding of Purified α-Protein to α and Non-α-DNA

The ^{32}P-labeled, 172 bp α-DNA monomer was mixed with homogeneous α-protein (fractions 17 and 18 from the hydroxyapatite column, see Figure 3E) and with increasing amounts of unlabeled E. coli competitor DNA (160 to 200 bp in length) before electrophoresis in a low-ionic-strength 4% polyacrylamide gel (see Experimental Procedures).

(A) Lanes a–j: α-DNA monomer (25,000 cpm; 2.5 ng of DNA) in the presence of α-protein (~0.5 ng), together with 0, 2.5, 5, 10, 20, 40, 80, 150, 300, and 600 ng of E. coli DNA per assay, respectively.

(B) Same as in A except that a ^{32}P-labeled, 180 bp DNA fragment excised from pBR322 with Msp I (see Experimental Procedures and Figure 1) was used instead of the α-DNA monomer at the same DNA amount per assay. Designations are as in Figure 1.

consistently protected in the DNAase footprints (Figure 5; M. Solomon, F. Strauss, and A. Varshavsky, unpublished data), suggesting that the terminal C–G pair is a necessary component of the ~7 bp α-protein binding site; however, this point remains to be demonstrated directly.

The footprinting technique could also be used to obtain a minimum estimate of the affinity of α-protein for α-DNA. As discussed previously (Brent and Ptashne, 1981), the apparent equilibrium association constant is equal to the reciprocal of the molar concentration of the unbound form of a DNA-binding species (assumed to be a 10,000 molecular weight α-protein monomer) at which the relative footprint intensity is half-maximal. Under conditions of the assay shown in Figure 5, the lowest α-protein concentration used (8 nM) produced a DNAase footprint pattern of almost the same relative intensity (sites I–III) as that seen with higher concentrations of α-protein (Figure 5). Thus the apparent affinity of α-protein for each of the three α-DNA-binding sites (assuming the absence of cooperative binding) is at least 1×10^8 M^{-1}; this figure is definitely an underestimate, since the DNA concentration in the assay (~1 nM in specific binding sites; see Figure 5) was not much smaller than the lowest concentration of α-protein used (8 nM), and since half-protection occurs at a lower α-protein concentration when less DNA is used (data not shown). Indeed, when the preformed α-protein–α-DNA monomer complex is progressively diluted with the binding buffer and the percentage of α-DNA still in the complex is determined by the gel electrophoresis assay (see Experimental Procedures), the equilibrium association constant is $\sim 5 \times 10^{10}$ M^{-1}, in agreement with the most recent footprinting results (M. Solomon, F. Strauss, and A. Varshavsky, unpublished data).

The DNAase footprint pattern did not change upon a 200-fold increase in the α-protein concentration (Figure 5 and data not shown), indicating that the affinity of α-protein for specific sites I–III on α-DNA is at least 200 times higher than its affinity for nonspecific α-DNA sites. This minimum estimate of the relative binding specificity is in agreement with the estimates made from the band-competition experiments of the type shown in Figures 1 and 4, indicating that α-protein binds to sites I–III on α-DNA approximately 10^3 times more tightly than to DNA outside these sites (Figures 1, 4, 5, and data not shown).

Examination of the mutual arrangement of sites I–III on α-DNA (Figures 5 and 6) has revealed striking spatial regularities, suggesting a nucleosome-positioning function for α-protein as discussed below.

Figure 3. Electrophoretic analysis of α-Protein in Nuclear Extracts and in Chromatographic Fractions

(A, B) Aliquots from even-numbered fractions (10–72) from the phosphocellulose column (see Experimental Procedures and Figure 2) were electrophoresed in 30 cm long, 18% polyacrylamide–SDS gels and stained with silver (see Experimental Procedures).

(C) An area framed in B is enlarged to show the distribution of α-protein-containing fractions relative to those containing HMG14 and HMG17.

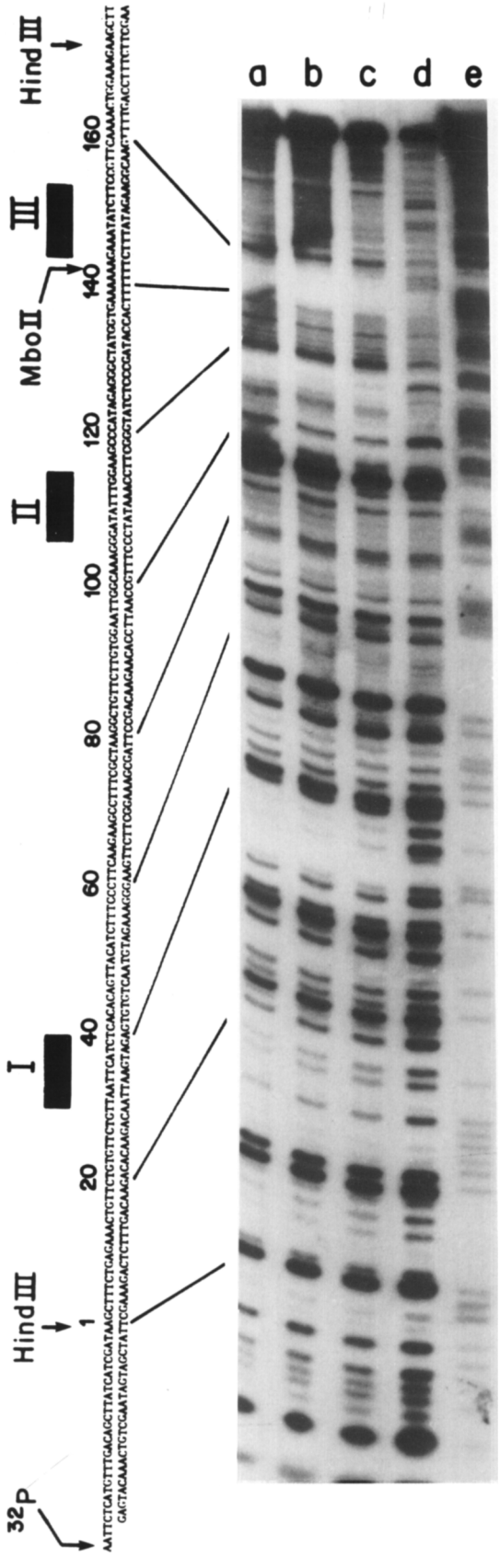
(D) Polyacrylamide–SDS electrophoretic patterns of the fractions from a DNA–Sephacryl column (the second purification step; see Figure 2C and Experimental Procedures).

(E) Polyacrylamide–SDS electrophoretic patterns of the fractions from a hydroxyapatite column (the third and last purification step; α-protein eluted at ~1.2 M NaCl; see Experimental Procedures).

(F) A 0.35 M NaCl extract of CV-1 nuclei (see Experimental Procedures) was made 5% (w/v) in CCl_3COOH; precipitated proteins were removed by centrifugation; proteins in the supernatant were precipitated with 25% (w/v) CCl_3COOH, redissolved, and subjected to the two-dimensional electrophoretic analysis (first dimension, acetic acid–urea; second dimension, SDS; see Experimental Procedures).

(G) A 0.2 N HCl extract of CV-1 nuclei was made 5% (w/v) in CCl_3COOH; soluble proteins were then processed for electrophoretic analysis as described in F. A small proportion of the core histones was recovered in the supernatant after treatment with 5% CCl_3COOH.

Designations: T (in A–E), total CV-1 nuclear protein soluble in SDS; E (in A–E), total protein in the 0.35 M NaCl extract of CV-1 nuclei; O (in E), the flowthrough fraction from the hydroxyapatite column; I (in E), proteins in the pooled DNA–Sephacryl fractions 28–33 (in D) before hydroxyapatite chromatography. Major histone species (H1, H2A, H2B, H3, H4), HMG14 and HMG17 proteins, and α-protein are indicated. Also shown are the positions of ubiquitin–H2A semihistone (uH2A) and of a protein, X (see Discussion), which almost comigrates with HMG14 in the SDS dimension (in F and G).



Discussion

The Band-Competition Assay as a General Method for Detecting Specific DNA-Binding Proteins in Crude Extracts

Although many of the current assays for DNA-binding proteins can be used to detect specific DNA–protein interactions in crude extracts, these assays are often not sufficiently sensitive or selective, particularly for DNA-binding proteins with a moderate degree of nucleotide sequence specificity. Since moderate sequence specificity was an expected property of a satellite-DNA-binding protein (see below), we devised a more sensitive, generally applicable approach, the band-competition assay (see Results), based upon an electrophoretic method for detecting complexes of DNA with purified DNA-binding proteins (Garner and Revzin, 1981; Fried and Crothers, 1981). The latter method is in turn based on the technique of low-ionic-strength gel electrophoresis developed originally for the analysis of oligo- and mononucleosomes (Varshavsky et al., 1976).

Using the band-competition assay, we have detected an α-DNA-binding protein (α-protein) in a crude CV-1 nuclear extract; another version of this assay, in which the concentration of competitor DNA was kept constant, provided a direct and sensitive means to monitor the purification of α-protein (see Results). Analogous assays with purified bacterial proteins of known sequence specificity and a nonspecific competitor DNA have been used recently for other purposes by Kolb et al. (1983) and by Fried and Crothers (1984).

Arrangement of α-Protein Sites on α-DNA Suggests a Specific α-Nucleosome Phasing Frame

As can be seen from Figures 5 and 6, α-protein binding sites II and III on α-DNA contain inverted 7 bp repeats of each other. The 7 bp "core" sequence of sites II and III (5′-AAATATC-3′) is neither a palindrome nor pure (A+T) DNA (the latter assumes that the length of an α-protein binding site is not less than 7 bp; although suggested by the data, this is not yet proven; see Results). Site I (5′-TTAATTC) is only weakly homologous (3 of 7 residues) to the 7 bp "core" sequence of sites II and III (5′-AAATATC; Figures 5

Figure 5. DNAase I Footprint of Purified α-Protein on α-DNA

The Eco RI–Hha I DNA fragment of pFS522, containing the cloned 172 bp α-DNA monomer (see Experimental Procedures), was end-labeled with ^{32}P at the Eco RI site as indicated. The fragment (1 ng) was mixed in 25 μl of 0.1% Triton X-100, 70 mM NaCl, 5 mM $MgCl_2$, 10 mM 2-mercaptoethanol, 10 mM Tris-HCl (pH 7.5), with 50 ng (lane a), 25 ng (lane b), 10 ng (lane c), and 0 ng (lane d) of the purified α-protein, and thereafter digested with 1 ng of DNAase I (Worthington) for 10 min at 37°C. Digestion was stopped with EDTA. DNA was purified, denatured, and electrophoresed in an 8% polyacrylamide–7 M urea gel as previously described (Maxam and Gilbert, 1980). The product of a Maxam–Gilbert "G+A" cleavage reaction of the same fragment was used as a marker (lane e). Position of the 172 bp α-DNA insert in the Eco RI–Hha I fragment is demarcated by the Hind III sites shown. Blocks I, II, and III span the regions of α-DNA bound by α-protein as detected by DNAase I footprinting.

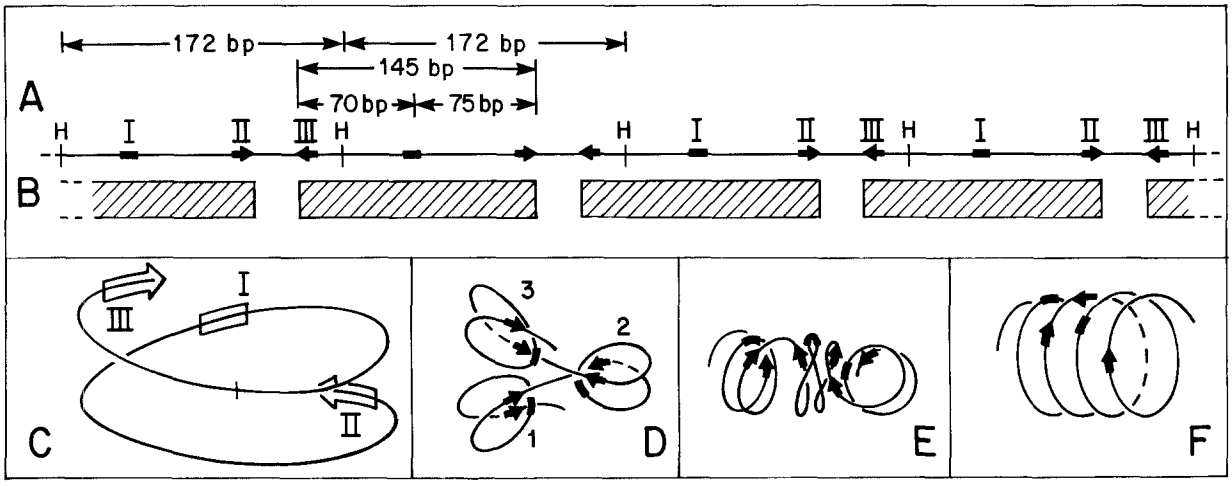


Figure 6. Distribution of α-Protein-Binding Sites on α-DNA: Possible Implications for α-Nucleosome Phasing and Packing

(A) Arrangement of α-protein-binding sites I, II, and III within the 172 bp α-DNA repeats. Four such repeats, demarcated by Hind III sites (H), are shown, together with the corresponding distances in base pairs. Apparent lengths of sites I–III (~7 bp) are drawn slightly larger than to scale, with arrows on sites II and III to indicate that they contain inverted repeats of each other.

(B) Specific α-nucleosome phasing frame suggested by the distribution of the sites I–III on α-DNA. Stippled rectangles correspond to the positions of the 145 bp α-nucleosome cores. This frame is identical with the preferred α-nucleosome phasing frame detected previously in isolated chromatin by totally different approaches (Zhang et al., 1983; Wu et al., 1983; see Discussion).

(C) Mutual proximity of sites I–III on the folded DNA of the α-nucleosome phased as shown in B.

(D–F) Highly schematic diagrams of spatial arrangements of the phased α-nucleosomes and of the corresponding sites I–III (black boxes and arrows) in three models of nucleosome packing. In D, the higher order fiber is a zigzag helical ribbon, in which the repeat unit is a dinucleosome and the path of internucleosomal (linker) DNA is strongly different from that of intranucleosomal (core) DNA (Lohr and Van Holde, 1979; Worcel et al., 1981). In the case of the phased α-nucleosomes shown in B, the linker DNA (27 bp long) is shorter (and the corresponding fiber is tighter) than would appear from the scheme in D. Recent computations (L. Ulanovsky and E. Trifonov, personal communication) carried out within the framework of the zigzag model suggest that linker lengths of 26 or 27 bp would result in a particularly compact nucleosome packing (also see text). In E, the higher order fiber is a solenoid with approximately six radially oriented nucleosomes per solenoid turn (Finch and Klug, 1976; McGhee et al., 1983; Mitra et al., 1984). For clarity, only the three nucleosomes on the front surface of the fiber are shown, and the relative length of spacer DNA is depicted as greater than it would be in case of the phased α-nucleosomes in B. According to the solenoid model, the spacer DNA is supercoiled about the helix that passes through the nucleosome centers. In F, the higher order fiber is a regular DNA superhelix in which the path of linker DNA is identical with that of the intranucleosomal (core) DNA.

and 6). This was the first indication that α-protein interacts preferentially with a family of short, (A+T)-rich DNA sequences, rather than with a single such sequence. More recent evidence suggests that the number of different short (~7 bp) DNA sequences bound by α-protein in vitro with comparable affinities is greater than two; in particular, the dispersed, ~300 bp long, primate repetitive DNA known as the Alu sequence (Schmid and Jelinek, 1982) was recently found to bind preferentially to α-protein, in spite of the absence of exact homologies to sites I and II (III) (F. Strauss and A. Varshavsky, unpublished data).

Closer examination of the spacings between sites I, II, and III on adjacent α-DNA repeats (Figure 6) revealed a striking pattern: first, the outward border of site III is 145 bp away from the outward border of the inversely oriented site II on the adjacent α-DNA repeat (Figure 6A); second, site I lies almost exactly midway between sites II and III (Figure 6A). The 145 bp distance between sites II and III on adjacent α-DNA repeats (Figure 6A) is identical with the length of core mononucleosomal DNA, while the 27 bp DNA stretch between sites II and III within the Hind III–defined α-DNA repeat (Figure 6A) can be interpreted as the linker DNA. The striking mutual arrangement of sites I, II, and III suggests a specific α-nucleosome phasing frame (Figure 6B), in which sites II and III would delimit the sequence-specific position of the core histone octamer. In this model, α-protein molecules bound at sites II and III would be located at the outer borders of the α-nucleosomal core particle in addition to, rather than instead of, core histones. Such an arrangement would be strikingly similar to the relative positions of in vitro nucleosome-bound proteins HMG14 and HMG17 (Mardian et al., 1980; Sandeen et al., 1980; Albanese and Weintraub, 1980; Schröter and Bode, 1982; Swerdlow and Varshavsky, 1983; Stein and Townsend, 1983), except that α-protein would be bound sequence-specifically, thereby establishing the specific α-nucleosome phasing frame (Figures 6A and 6B).

Remarkably, the α-nucleosome phasing frame suggested by the arrangement of sites I–III is identical with the preferred frame detected previously in isolated CV-1 chromatin by Zhang et al. (1983) and by Wu et al. (1983, this laboratory), using two different direct methods, and also by Musich et al. (1982), using an indirect method.

In preliminary experiments we have found that the pattern of in vitro α-protein binding to bulk CV-1 mononucleosomes is strikingly similar to that of HMG14/17-nucleosome binding (R. Pan, F. Strauss, and A. Varshavsky, unpublished data). Moreover, one of the three abundant, nucleosome-associated HMG proteins seen by us previously in two-dimensional electrophoretic analyses of isolated CV-1

nucleosomes was later identified as α-protein; the other two proteins were HMG14 and HMG17 (Figure 7 in Wu et al., 1983, and data not shown). While both in vitro and in vivo α-nucleosome–α-protein reconstitution experiments would be necessary to address directly the postulated role for α-protein in organizing α-nucleosome phasing, the remarkable identity of the experimentally established preferred frame (Zhang et al., 1983; Wu et al., 1983) with the frame suggested by the arrangement of α-protein binding sites on α-DNA (Figures 6A and 6B) strongly suggests a causative role for α-protein in the formation and maintenance of the preferred α-nucleosome phasing frame. It remains to be determined whether the presence of minor α-nucleosome phasing frames in isolated chromatin (Zhang et al., 1983) is due to in vivo α-nucleosome arrangements not accounted for by the presence of α-protein, or whether at least some of the minor frames are the result of in vitro rearrangements caused by experimental manipulation.

Binding Sites I–III Would Be Brought into Mutual Proximity by DNA Folding in the Nucleosome

As can be seen from Figure 6C, bending of DNA in the phased α-nucleosome would bring the inversely oriented sites II and III into mutual proximity at one side of the nucleosome, close to the DNA "exit" point. Remarkably, site I would be located on the same "exit" side of the nucleosome, directly between sites II and III (Figure 6C). Both high symmetry of the arrangement of sites I–III on the folded α-nucleosomal DNA and their mutual proximity (Figure 6C) are consistent with a number of structural models for α-protein–α-nucleosome interactions, including cooperative binding of α-protein to sites I–III in α-nucleosomes. It also remains to be established whether α-protein binds individually to sites I–III on the folded nucleosomal DNA, or whether any combination of sites I–III forms a "compound" recognition site by virtue of their mutual proximity in the nucleosome (Figure 6C). The suggested cooperativity of binding of α-protein to "properly" spaced sites I–III in the α-nucleosome (Figure 6) may be crucial not only for the nucleosome-positioning function of α-protein per se but also for assuring preferential binding of α-protein to α-heterochromatin as compared with numerous, but less regularly spaced, homologs of sites I–III in the rest of the genome.

Figures 6D–6F show highly schematic diagrams of spatial arrangements of the phased α-nucleosomes and of the corresponding sites I–III in three models of the first order nucleosome packing ("zigzag," "solenoid," and "regular superhelix" models; see the legend to Figure 6 for references and further details). Nothing is known about the postulated role of α-protein in generating the compact higher order structure of constitutive α-heterochromatin; one possibility is that the α-protein-mediated, sequence-specific (phased) arrangement of α-nucleosomes (Figure 6B) by its very regularity allows formation of an especially compact and ordered nucleosome packing that would be described on the cytological level as heterochromatin. Other proteins, such as histone H1, may be required to bring about this higher order packing, with α-protein molecules possibly providing necessary nucleotide sequence specificity.

α-Protein May be a Phosphorylated HMG Protein That Is Additionally Phosphorylated upon Mitosis

Our survey of the literature on HMG proteins published before 1983 did not turn up any evidence that α-protein had been identified previously. However, in a recent work, Lund et al. (1983) described specific HMG proteins in human (HeLa) cells that they tentatively called HMG-I and HMG-M. Both the electrophoretic properties and amino acid composition of HMG-I (Lund et al., 1983) are indistinguishable from those of α-protein (Figures 3F and 3G and data not shown). Both HMG-I and HMG-M are phosphoproteins; HMG-I is much more abundant than HMG-M in interphase cells, while the reverse is true in metaphase cells (Lund et al., 1983). These data, taken together with the identity of amino acid compositions of HMG-I and HMG-M and their distinct phosphorylation patterns, strongly suggest that HMG-M is an additionally phosphorylated derivative of HMG-I, which is itself a phosphoprotein. Not only is α-protein apparently identical with HMG-I (see above) but, in addition, the relative position of the unidentified protein X in Figure 3G is indistinguishable from the position of HMG-M reported by Lund et al. (1983). Thus, although not proven, it is likely that protein X (Figure 3G) is derived from α-protein by an additional, mitosis-specific phosphorylation.

A Moderately Sequence-Specific Protein Can Be Bound In Vivo to Cognate DNA Sites if the Latter Are Abundant Enough

Procaryotic regulatory proteins bind their cognate DNA sequences with many orders of magnitude higher affinities than unrelated DNA; high specificity is required for these proteins to locate rare cognate DNA sequences in vivo (reviewed by von Hippel, 1979). It may therefore seem that a relatively moderate sequence specificity, such as that of α-protein for α-DNA, might not be relevant to its in vivo function. Simple computations of coupled equilibria show, on the contrary, that a considerable proportion of a moderately sequence-specific protein can be bound to its cognate DNA sequences in vivo, provided the cognate sequences are sufficiently abundant. In the approximation below we assume all specific DNA sites (D_s) indistinguishable in their high affinity for α-protein (α), and all nonspecific sites (D_n) identical in their low affinity.

$$\alpha + D_s \rightleftharpoons \alpha D_s \quad (1)$$

$$\alpha + D_n \rightleftharpoons \alpha D_n \quad (2)$$

Equilibrium association constants for reactions (1) and (2) are, respectively,

$$K_s = \frac{[\alpha D_s]}{[\alpha][D_s]} \quad (3)$$

$$K_n = \frac{[\alpha D_n]}{[\alpha][D_n]} \quad (4)$$

The corresponding conservation constraints are

$$[D_s^o] = [D_s] + [\alpha D_s] \quad (5)$$

$$[D_n^o] = [D_n] + [\alpha D_n] \quad (6)$$

$$[\alpha^o] = [\alpha] + [\alpha D_n] + [\alpha D_s] \quad (7)$$

where $[D_s^o]$, $[D_n^o]$ and $[\alpha^o]$ are total concentrations of specific DNA sites, nonspecific DNA sites, and α-protein, respectively.

In α-DNA there are three specific sites (I–III) per 172 bp repeat (see Results); thus the number of nonspecific α-DNA sites is 169, since each of the 172 base pairs except 3 is a beginning of a nonspecific site. Assuming further that non-α-DNA contains no α-protein-binding sites (definitely an underestimate, see above), and taking into account the weight content of α-DNA in the green monkey genome (~20%; Singer, 1982), we obtain the following conservative (maximum) estimate for the ratio of nonspecific to specific DNA-binding sites for α-protein in the green monkey genome:

$$\frac{[D_n^o]}{[D_s^o]} \approx \frac{169}{3} \cdot \frac{5}{1} \approx 280 \quad (8)$$

Suppose that a majority, for example 60%, of all bound α-protein molecules in the cell is bound specifically, i.e. that

$$[\alpha D_s] = 1.5[\alpha D_n] \quad (9)$$

Suppose, furthermore, that most, for example 90%, of specific DNA sites are occupied by α-protein in vivo, i.e., that

$$[\alpha D_s] = 9[D_s] \quad (10)$$

Substituting (8), (9), and (10) into equations (3)–(7) yields, after straightforward algebraic manipulations, the following value for the ratio K_s/K_n

$$\frac{K_s}{K_n} = \frac{[\alpha D_s]\,[D_n]}{[\alpha D_n]\,[D_s]} \approx 4191 \approx 4 \times 10^3 \quad (11)$$

Thus a less than 10^4-fold preference for specific versus nonspecific DNA sites, with specific sites present at the relative abundance of sites I–III in the green monkey genome, would result in the majority of bound α-protein molecules being bound specifically. The calculated degree of sequence specificity required is not inconsistent with the current approximate estimate of the nucleotide sequence specificity of α-protein (~10^3-fold preference for specific versus nonspecific sites on α-DNA; see Results). The above 60% figure for specifically bound α-protein should be compared with those for bacterial regulatory proteins that recognize one or a few specific DNA sites per bacterial genome; for instance, less than 10% of DNA-bound *lac* repressor is bound specifically in vivo, in spite of its ~10^8-fold preference for cognate DNA sites versus nonspecific DNA sites (von Hippel, 1979).

The important conclusion is that given sufficient abundance of specific DNA sites, a relatively low degree of nucleotide sequence specificity is required for a protein to be bound specifically in vivo. This suggests that not only α-protein but also other abundant chromosomal proteins, such as, for example, HMG14 and HMG17, may function via sequence-specific DNA recognition. In view of the above argument, the weak but detectable preferences of histones themselves for apparently specific and abundant short sequences in DNA (Simpson and Stafford, 1983; Zhurkin, 1983) may also have a functional significance; for instance, by allowing formation of weakly phased nucleosome arrays in the absence of nucleosome-positioning proteins or by influencing the choice of a specific phasing frame by a nucleosome-positioning protein in a region where alternative arrangements are possible.

In the case of α-protein, the minimum number of specific binding sites in the green monkey genome equals the number of sites I–III in α-DNA (~3×10^6 sites per haploid genome; see above). The number of specific α-protein binding sites outside of α-DNA domains, while quite possibly comparable with that of sites I–III in α-DNA, cannot be reliably estimated at present for two reasons. First, it is clear that α-protein recognizes a *family* of short (~7 bp), (A+T)-rich, related sequences, and that this family consists of more than two different members (see above). Second, the striking mutual arrangement of sites I–III, which defines the α-nucleosome phasing frame (Figure 6 and the discussion above), suggests that a "proper" spacing of α-protein binding sites on DNA may be crucial for high-affinity (cooperative) binding of α-protein to its cognate DNA sites in vivo. In view of this uncertainty, accurate determination of α-protein distribution within (versus outside) α-heterochromatin in vivo would require α-protein-specific antibodies and immunocytochemical approaches; this work is in progress.

Possible Functions of α-Protein Inside and Outside of α-DNA Domains: The Notion of a Nucleosome-Positioning Protein

While there appears to be a significant random component in the distribution of bulk nucleosomes relative to the total genomic DNA sequence (reviewed by Kornberg, 1981), it is becoming increasingly clear that both isolated and in vitro formed nucleosomes can be arranged nonrandomly with regard to the corresponding nucleotide sequences in specific subsets of the genome (Varshavsky et al., 1979; Wasylyk et al., 1979; Igo-Kemenes et al., 1980; Gottesfeld and Bloomer, 1980; Bryan et al., 1981; Bloom and Carbon, 1982; Wittig and Wittig, 1982; Nasmyth, 1982; Musich et al., 1982; Worcell et al., 1983; Young and Carroll, 1983; Pratt and Hattman, 1983; Zhang et al., 1983; Wu et al., 1983; Gottschling et al., 1983; see also Keene and Elgin, 1983). The mechanisms of formation and maintenance of apparently sequence-specific (phased) nucleosome arrangements remain unclear; they may involve intrinsic binding preferences of octameric histone cores for specific sequences in DNA (Chao et al., 1979; Trifonov and Sussman, 1980; Simpson and Stafford, 1983; Zhurkin, 1983)

or generation of statistically nonrandom arrangements of nucleosomes in the vicinities of sequence-specific "start points" provided by relatively rare, highly sequence-specific DNA-binding proteins (Kornberg, 1981).

Our findings with α-protein (see Results) suggest a new mechanism, in which a phased arrangement of nucleosomes on a tandemly repetitive DNA is achieved by stoichiometric binding of a relatively abundant *nucleosome-positioning* protein, such as α-protein, to regularly and appropriately spaced cognate DNA sequences in the repetitive DNA. The sequence-specifically bound nucleosome-positioning protein would then direct formation of a phased nucleosome array by interacting specifically with nucleosomal histones and by providing properly spaced DNA sites for nucleosome assembly. Cooperativity of the above interactions due to a regular arrangement of cognate DNA sites for a nucleosome-positioning protein may be a crucial factor in driving the assembly of a phased nucleosome array. It is likely that for a phased nucleosome array to be formed, α-protein should be bound to a majority, but not necessarily to all, of the cognate DNA sequences within α-DNA, because each α-protein-positioned nucleosome may be formally viewed as a sequence-specific nucleosome assembly start point, so that nucleosomes assembled in the immediate vicinity of a positioned nucleosome would be at least partially phased as a result of statistical effects discussed by Kornberg (1981). Occasional irregularities in the repeat length and quasi-random deviations from the consensus sequence that are characteristic for all satellite DNAs (Singer, 1982), therefore, would not result in large-scale dislocations of the predominant α-nucleosome phasing frame mediated by α-protein.

According to this model, α-protein may underlie both the compactness and transcriptional inactivity of constitutive α-heterochromatin by producing a crystal-like, phased arrangement of α-nucleosomes, which by its very regularity may allow formation of an especially compact and ordered nucleosome packing. Whether α-protein is sufficient to bring about such transitions, or whether it functions by providing nucleotide-sequence-specific sites for interactions with other proteins, such as histone H1, remains to be established.

It is possible that there exists more than one type of a nucleosome-positioning, or phasing, protein, each type recognizing different families of short nucleotide sequences. Specifically, we suggest that the existence of a phasing protein which recognizes multiple, chemically modified (methylated?) sites in mammalian X-chromosomal DNA may be sufficient to account for major features of the phenomenon of X-chromosome inactivation (reviewed by Martin, 1982). More generally, we suggest a temporally controlled, phasing-protein-mediated positioning of nucleosomes at or near DNA sites required for initiation of transcription as a new genetic regulatory mechanism. For instance, an α-protein-like phasing protein that becomes active late in amphibian (Xenopus) embryogenesis and binds to specific DNA sequences near oocyte-type, but not somatic-type, 5S RNA genes may be required to account for the selective repression of oocyte-type 5S rDNA transcription observed in late Xenopus embryos and in adult somatic cells (reviewed by Korn, 1982). These speculations assume that phasing proteins may also function outside of heterochromatin by phasing short arrays of nucleosomes on repetitive or nonrepetitive DNA. If so, transitions in specific chromosomal domains mediated by phasing proteins may underlie the known but poorly understood ability of eucaryotic cells to repress stably or to activate specific genes outside of constitutive heterochromatin, and to propagate such regulatory patterns through successive generations of cells in development and differentiation.

Experimental Procedures

Cell Culture

African green monkey cells (CV-1 line) were maintained as monolayers in 14 cm plastic plates (Nunc) in Eagle's MEM medium supplemented with penicillin/streptomycin and 10% calf serum (GIBCO). For a large-scale purification of α-protein, CV-1 cells were grown to confluence in forty 850 cm^2 roller bottles (Falcon) to yield approximately 5×10^9 cells.

Preparation of CV-1 Nuclei

Confluent CV-1 monolayers were rinsed with cold 0.1 M NaCl, 50 mM KCl, 10 mM Na-phosphate (pH 7.2), scraped with a rubber policeman into a small volume of the same buffer, and centrifuged at 500 × g for 3 min. The cell pellet was resuspended, washed once, and resuspended again in 3 volumes of 0.23 M sucrose in buffer A (60 mM KCl, 15 mM NaCl, 0.25 mM $MgCl_2$, 0.5 mM Na-EGTA, 0.5 mM spermine, 0.15 mM spermidine, 14 mM 2-mercaptoethanol, 15 mM Tris-HCl [pH 7.4], containing freshly added proteinase inhibitors phenylmethylsulfonyl fluoride [PMSF, 0.2 mM], and 5 μg/ml each of antipain, leupeptin, chymostatin, and pepstatin A [Sigma]). Cells were disrupted in a motor-driven Potter-Elvehjem homogenizer (3000 rpm, ~12 strokes); phase-contrast microscopy was used to determine the optimal number of strokes. The lysate (60 ml, $\sim 5 \times 10^9$ nuclei) was diluted 3-fold with 2.0 M sucrose in buffer A, layered onto a 9 ml cushion of 1.7 M sucrose in buffer A, and centrifuged in the HB-4 rotor at 13,000 rpm for 45 min at 4°C in the RC-5B centrifuge (Sorvall). The nuclei were resuspended, washed once, and resuspended again in 0.23 M sucrose in buffer A to a final DNA concentration of ~2 mg/ml. The nuclei were then either used immediately or stored at −70°C, without noticeable changes in the results.

Extraction of α-Protein from CV-1 Nuclei

Purified CV-1 nuclei were pelleted by low-speed centrifugation and resuspended by vortexing to a final DNA concentration of ~0.4 mg/ml in 0.35 M NaCl, 5 mM Na-EDTA, 10 mM 2-mercaptoethanol, 10 mM Tris-HCl (pH 7.5) containing the five proteinase inhibitors (PMSF, antipain, leupeptin, chymostatin, and pepstatin A) at the concentrations used in the nuclei isolation. After 30 min at 0°C with occasional vortexing, the suspension was centrifuged at 10,000 × g for 15 min. The supernatant containing α-protein was used either immediately or after storage at −70°C in the presence of 15% glycerol. Extractions of nuclei with either 0.6 M NaCl or 2 M NaCl did not result in significantly higher yields of α-DNA-binding activity.

Preparation of α-Satellite DNA Probes

For the band-competition experiments (see Figures 1, 2, and 4), the 172 bp α-DNA monomer was isolated following complete digestion of purified CV-1 nuclear DNA (Wu et al., 1983) with either Hind III or Mbo II (see Figure 5). The two α-DNA monomers were purified separately by preparative electrophoresis in 6% polyacrylamide, eluted from the gels, and thereafter separately end-labeled with ^{32}P by T4 polynucleotide kinase (New England Biolabs) and γ-^{32}P-ATP (~5000 Ci/mmole, Amersham), following a treatment with bacterial alkaline phosphatase (Bethesda Research Laboratories) as described previously (Maxam and Gilbert, 1980). The labeled DNA was purified by extractions with $CHCl_3$:isopropanol (24:1) in the presence of 1% SDS, 1 M NaCl, and ethanol-precipitated in the presence of 10 μg/ml of

linear polyacrylamide as a carrier (F. Strauss, unpublished technique). Approximately equal ^{32}P-counts of the Hind III– and Mbo II–produced α-DNA monomers were then mixed together to yield the final DNA sample. This precaution was taken to make it less likely that a specific restriction site used to generate the α-DNA monomer would span a binding site for an α-DNA-binding protein. In later experiments, the use of the Hind III–produced α-DNA monomer alone gave the same results.

For DNAase I footprinting experiments, a plasmid (pFS522) containing an α-DNA monomer was constructed by inserting the Hind III–produced α-DNA monomer (see above) into the Hind III site of pBR322, using standard cloning methods (Maniatis et al., 1982). The pFS522 plasmid was cut at the single Eco RI site, end-labeled with ^{32}P (see above), then recut with Hha I, and the 273 bp Eco RI–Hha I fragment labeled at the Eco RI site and containing the α-DNA insert was purified by polyacrylamide gel electrophoresis. The nucleotide sequence of the cloned α-DNA insert in the Eco RI–Hha I fragment (see Figure 5) was determined by the Maxam–Gilbert method (Maxam and Gilbert, 1980), and was found to differ from the consensus α-DNA sequence (Rosenberg et al., 1978) in 3 of 172 positions: 18 (C→T), 45 (G→C), and 143 (G→A) (data not shown).

Preparation of E. coli Competitor DNA

Purified E. coli DNA was sonicated to an average chain length of ~1 kb, ethanol-precipitated, dissolved in 1 mM Na-EDTA, 10 mM Tris-HCl (pH 7.5), and dialyzed against the same buffer. In some of the experiments shorter DNA fragments were used as a competitor; they were produced by partially digesting E. coli DNA with staphylococcal nuclease, fractionating the digest by electrophoresis in a 6% polyacrylamide gel, and eluting DNA fragments of 160 to 200 bp in length. No significant difference in competing efficiency was seen between these two E. coli DNA preparations when used at equal DNA concentrations (data not shown).

Detection of α-Protein in Crude Extracts by the Band-Competition Assay

The ^{32}P-labeled, 172 bp α-DNA monomer, unlabeled E. coli DNA competitor, and nuclear extract were mixed in 25 μl of 0.1% Triton X-100, 4% glycerol, 1 mM Na-EDTA, 10 mM 2-mercaptoethanol, 10 mM Tris-HCl (pH 7.5) (final concentrations) containing NaCl, so that the final NaCl concentration (or the $[Na^+ + K^+]$ concentration in other experiments) was 80 ± 10 mM. Nuclear extract (1 to 5 μl) was added last. The mixture was incubated for 30 min at room temperature and thereafter loaded onto a low-ionic-strength 4% polyacrylamide gel (acrylamide:bisacrylamide weight ratio of 30:1) containing 1 mM Na-EDTA, 3.3 mM Na-acetate, 6.7 mM Tris (pH 7.5). The gel (0.15 × 16 cm) was preelectrophoresed for 2 hr at ~12 V/cm. Electrophoresis was carried out at the same voltage gradient (200 V for a 16 cm gel) for 4 hr at 4°C with the buffer stirred in both compartments and recirculated between the compartments (Varshavsky et al., 1976; Strauss and Prunell, 1982). The gel was then soaked in 5% glycerol, dried, and autoradiographed.

While the presence of Triton X-100 in the binding buffer (see above) was unnecessary for crude extracts, it was necessary to prevent losses of the purified α-protein in analogous assays (Figure 4 and data not shown). Either the presence of up to 10 mM $MgCl_2$ (in the absence of EDTA) in the binding buffer or the presence of 0.5 mM $MgCl_2$ (in the presence of 0.5 mM EGTA) in the gel buffer did not significantly influence the results (data not shown).

Purification of α-Protein

Phosphocellulose Chromatography

The 0.35 M NaCl extract (~120 ml) was dialyzed for 6 hr (with one change of the buffer) against 500 ml of 1 mM Na-EDTA, 10 mM 2-mercaptoethanol, 10 mM Tris-HCl (pH 7.5), containing the five proteinase inhibitors (PMSF, antipain, leupeptin, chymostatin, and pepstatin A) at the concentrations used in the nuclei isolation (see above). A precipitate appeared during dialysis and was discarded by centrifugation. The dialyzed extract was loaded onto a 21 ml phosphocellulose column (P11, Whatman) preequilibrated with 1 mM Na-EDTA, 10 mM 2-mercaptoethanol, 10 mM K-phosphate (pH 7.5). The column was then washed with 50 ml of the same buffer followed by elution with a 400 ml linear gradient of K-phosphate, pH 7.5 (from 0.01 to 1 M), in 1 mM Na-EDTA, 10 mM 2-mercaptoethanol. The concentration of K-phosphate in the collected 4 ml fractions was determined by refractometry.

DNA–Sephacryl Chromatography

The α-protein-containing fractions 49–55 from the phosphocellulose column (see Figures 2A, 2B, 3A–3C) were pooled, dialyzed against 1 liter of 10 mM NaCl, 1 mM Na-EDTA, 10 mM 2-mercaptoethanol, 10 mM Tris-HCl (pH 7.5) for 7 hr with two changes of the buffer, and thereafter loaded onto a 2 ml DNA–Sephacryl column preequilibrated with the same buffer. The DNA–Sephacryl (2 mg of DNA per ml of gel) was prepared by covalently linking purified double-stranded E. coli DNA (average chain length of approximately 1 kb; see above) to Sephacryl S-500 (Pharmacia), using the cyanogen bromide procedure of Büneman et al. (1982). After washing the column with the equilibration buffer, proteins were eluted with a 20 ml linear gradient of NaCl (from 0.02 to 0.5 M) in 1 mM Na-EDTA, 10 mM 2-mercaptoethanol, 10 mM Tris-HCl (pH 7.5), and 0.5 ml fractions were collected.

Hydroxyapatite Chromatography

The α-protein-containing fractions 28–33 (see Figures 2C and 3D) from the DNA–Sephacryl column were pooled, diluted 3-fold with H_2O, and loaded onto a 0.5 ml hydroxyapatite column (BioRad) preequilibrated with 0.1 M NaCl, 10 mM 2-mercaptoethanol, 10 mM K-phosphate (pH 7.5). After washing the column with the same buffer, proteins were eluted with a 25 ml linear gradient of NaCl (from 0.1 M to 3 M) in 10 mM 2-mercaptoethanol and 10 mM K-phosphate (pH 7.5), and 0.6 ml fractions were collected. Fractions 17 and 18 from the hydroxyapatite column, containing the α-DNA-binding activity and the purified α-protein (see Figure 3E), were pooled, dialyzed against 20% glycerol, 0.35 M NaCl, 1 mM Na-EDTA, 10 mM 2-mercaptoethanol, 10 mM Tris HCl (pH 7.5), and stored in small aliquots at −70°C. Approximately 200 μg of pure α-protein was obtained from 5 × 10^9 CV-1 cells.

Gel Electrophoresis of Proteins

Proteins were electrophoresed on discontinuous polyacrylamide–SDS gels (30 cm long, 0.15 cm thick; 18% separating gel) according to Thomas and Kornberg (1975), except that 2-mercaptoethanol was omitted from the sample buffer to avoid artifacts upon silver staining (unpublished data). For two-dimensional protein separations (see Figures 3F and 3G), the first-dimension electrophoresis was carried out in the 15% polyacrylamide, acetic acid, urea system of Panyim and Chalkley (1969). The first-dimension gel slice was then soaked in the SDS sample buffer for 1 hr at 37°C and loaded onto the 30 cm discontinuous SDS gel described above. Silver staining after SDS gel electrophoresis was carried out essentially as described by Oakley et al. (1980) except that gels were first soaked in 50% CH_3OH for 3–4 hr with several changes of 50% CH_3OH, and after soaking in ammoniacal silver solution, gels were subjected to three 10 min washes in the same solution diluted 200-fold. These modifications greatly reduced the background staining.

Acknowledgments

We are greatly indebted to Mark Solomon and Louis Levinger for many helpful discussions and advice, and to Daniel Finley and Mark Solomon for their comments on the manuscript. We also thank John Smart (Biogen) for determination of the amino acid composition of α-protein and Ron Rodriguez for his help with protein fractionation. This work was supported by grants to A. V. from the National Cancer Institute (CA30367) and from the National Institute of General Medical Sciences (GM31530). F. S. was supported by the Centre National de la Recherche Scientifique (France) and by a Fogarty International Fellowship from the National Institutes of Health.

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

Received April 4, 1984; revised April 24, 1984

References

Albanese, I., and Weintraub, H. (1980). Electrophoretic separation of a class of nucleosomes enriched in HMG14 and 17 and actively transcribed globin genes. Nucl. Acids Res. *8*, 2787–2805.

Avilla, J., de Garcini, E. M., Wandosell, F., Villasante, A., Sogo, J. M., and Villanueva, N. (1983). Microtubule-associated protein MAP_2 preferentially binds to dA/dT sequence present in mouse satellite DNA. EMBO J. *2,* 1223–1234.

Beerman, M., and Meyer, G. F. (1980). Chromatin rings as products of chromatin diminution in cyclops. Chromosoma *77,* 277–283.

Bennet, F. C., Rosenfeld, B. I., Huang, C. K., and Yeoman, L. C. (1982). Evidence for two conformational forms of nonhistone protein BA which differ in their affinity for DNA. Biochem. Biophys. Res. Communs. *104,* 649–656.

Bloom, K. S., and Carbon, J. (1982). Yeast centromere DNA is in a unique and highly ordered structure in chromosomes and small circular minichromosomes. Cell *29,* 305–317.

Brent, R., and Ptashne, M. (1981). Mechanism of action of the lexA gene product. Proc. Nat. Acad Sci. USA *78,* 4204–4208.

Brutlag, D. (1980). Molecular arrangement and evolution of heterochromatic DNA. Ann. Rev. Genet. *14,* 121–144.

Bryan, P. N., Hofstetter, H., and Birnstiel, M. L. (1981). Nucleosome arrangement on tRNA genes of Xenopus laevis. Cell *27,* 459–466.

Büneman, H., Westhoff, P., and Herrmann, R. G. (1982). Immobilization of denatured DNA to macroporous supports. Nucl. Acids Res. *10,* 7163–7180.

Cavalier-Smith, T. (1982). Skeletal DNA and evolution of genome size. Ann. Rev. Biophys. Bioeng. *11,* 273–302.

Chao, M. V., Gralla, J., and Martinson, H. G. (1979). DNA sequence directs placement of histone cores on restriction fragments during nucleosome formation. Biochemistry *8,* 1068–1074.

Finch, J. T., and Klug, A. (1976). Solenoidal model for superstructure in chromatin. Proc. Nat. Acad. Sci. USA *73,* 1898–1902.

Fried, M. G., and Crothers, D. M. (1981) Equilibria and kinetics of lac repressor–operator interactions by polyacrylamide gel electrophoresis. Nucl. Acids Res. *9,* 6505–6525.

Fried, M. G., and Crothers, D. M. (1984). Equilibrium studies of the cAMP receptor protein–DNA interaction. J. Mol. Biol. *172,* 241–262.

Galas, D., and Schmitz, A. (1978). DNase footprinting. Nucl. Acids Res. *5,* 3157–3170.

Garner, M. M., and Revzin, A. (1981). A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions. Nucl. Acids Res. *9,* 3047–3060.

Garreau, H., and Williams, J. G. (1983). Two nuclear DNA-binding proteins of Dictyostelium discoideum with a high affinity for poly(dA)–poly(dT). Nucl. Acids Res. *11,* 8473–8484.

Gatty, M., Smith, D. A., and Baker, B. S. (1983). A gene controlling condensation of heterochromatin in Drosophila. Science *221,* 83–85.

Gottesfeld, J. M., and Bloomer, L. S. (1980). Nonrandom alignment of nucleosomes on 5S RNA genes of X. laevis. Cell *21,* 751–760.

Gottschling, D. E., Palen, T. E., and Cech, T. R. (1983). Different nucleosome spacing in transcribed and non-transcribed regions of the ribosomal RNA gene. Nucl. Acids Res. *11,* 2093–2109.

Hilliker, A. J., Appels, R., and Schalet, A. (1980). The genetic analysis of D. melanogaster heterochromatin. Cell *21,* 607–619.

Hsieh, T., and Brutlag, D. (1978). A protein that preferentially binds Drosophila satellite DNA. Proc. Nat. Acad. Sci. USA *76,* 726–730.

Igo-Kemenes, T., Omori, A., and Zachau, H. G. (1980). Nonrandom arrangement of nucleosomes in satellite I–containing chromatin of rat liver. Nucl. Acids Res. *8,* 5377–5390.

John, B., and Miklos, G. L. C. (1979). Functional aspects of satellite DNA and heterochromatin. Int. Rev. Cytol. *58,* 1–113.

Keene, M. A., and Elgin, S. C. R. (1983). Patterns of DNA structural polymorphism and their evolutionary implications. Cell *36,* 121–129.

Klobutcher, L. A., Swanton, M. T., Donini, P., and Prescott, D. M. (1981). All gene-sized DNA molecules in four species of hypotrichs have the same terminal sequence and an unusual 3′ terminus. Proc. Nat. Acad. Sci. USA *78,* 3015–3019.

Kolb, A., Spassky, A., Chapon, C., Blazy, B., and Buc, H. (1983). On the different binding affinities of CRP at the lac, gal and malT promoter regions. Nucl. Acids Res. *11,* 7833–7852.

Korn, L. J. (1982). Transcription of Xenopus 5S ribosomal RNA genes. Nature *295,* 101–105.

Kornberg, R. D. (1981). Location of nucleosomes in chromatin: specific or statistical? Nature *292,* 579–580.

Levinger, L., and Varshavsky, A. (1982a). Selective arrangement of ubiquitinated and D1 protein-containing nucleosomes within the Drosophila genome. Cell *28,* 375–385.

Levinger, L., and Varshavsky, A. (1982b). Protein D1 preferentially binds (A+T)-rich DNA in vitro and is a component of Drosophila melanogaster nucleosomes containing (A+T)-rich satellite DNA. Proc. Nat. Acad. Sci. USA *79,* 7152–7156.

Levinger, L., Barsoum, J., and Varshavsky, A. (1981). Two-dimensional hybridization mapping of nucleosomes: comparison of DNA and protein patterns. J. Mol. Biol. *146,* 287–304.

Lohr, D., and Van Holde, K. (1979). Organization of spacer DNA in chromatin. Proc. Nat. Acad. Sci. USA *76,* 6326–6330.

Lund, T., Holtlund, J., Fredriksen, M., and Laland, S. G. (1983). On the presence of two new HMG proteins in HeLa S3 cells. FEBS Lett. *152,* 163–167.

Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982). Molecular Cloning. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory).

Mardian, J. K. W., Paton, A. E., Bunick, G. J., and Olins, D. E. (1980). Nucleosome cores have two specific binding sites for nonhistone chromosomal proteins HMG14 and HMG17. Science *209,* 1534–1536.

Martin, G. R. (1982). X-chromosome inactivation in mammals. Cell *29,* 721–724.

Maxam, A., and Gilbert, W. (1980). Sequencing end-labeled DNA with base-specific chemical cleavages. Meth. Enzymol. *65,* 499–525.

McGhee, J. D., Nickol, J. M., Felsenfeld, G., and Rau, D. C. (1983). Higher order structure of chromatin: orientation of nucleosomes within the 30 nm chromatin solenoid is independent of species and spacer length. Cell *33,* 831–841.

Mitra, S., Sen, D., and Crothers, D. M. (1984). Orientation of nucleosomes and linker DNA in calf thymus chromatin determined by photochemical dichroism. Nature *308,* 247–250.

Müller, F., Walker, P., Alby, P., Neuhaus, H., Felder, H., Back, E., and Tobbler, H. (1983). Nucleotide sequence of the satellite DNA contained in the eliminated genome of Ascaris lumbricoides. Nucl. Acids Res. *10,* 7493–7510.

Musich, P. R., Brown, F. L., and Maio, J. J. (1982). Nucleosome phasing and micrococcal nuclease cleavage of African green monkey component α DNA. Proc. Nat. Acad. Sci. USA *79,* 118–122.

Nasmyth, K. A. (1982). The regulation of yeast mating-type chromatin structure by *SIR:* an action at a distance affecting both transcription and transposition. Cell *30,* 567–578.

Oakley, B. R., Kirsch, D. R., and Morris, N. R. (1980). A simplified ultrasensitive silver stain for detecting proteins in polyacrylamide gels. Analyt. Biochem. *105,* 361–368.

Panyim, S., and Chalkley, R. (1969). High-resolution acrylamide gel electrophoresis of histones. Arch. Biochem. Biophys. *130,* 337–346.

Peacock, W. J., Lohe, A. R., Gerlach, W. L., Dunsmuir, P., Dennis, E. S., and Appels, R. (1978). Fine structure and evolution of DNA in heterochromatin. Cold Spring Harbor Symp. Quant. Biol. *42,* 1121–1136.

Pratt, K., and Hattman, S. (1983). Nucleosome phasing in Tetrahymena macronuclei. J. Protozool. *30,* 592–598.

Rodriguez-Alfageme, C. R., Rudkin, G. T., and Cohen, L. H. (1980). Isolation, properties and cellular distribution of D1, a chromosomal protein of Drosophila. Chromosoma *78,* 1–31.

Rosenberg, H., Singer, M., and Rosenberg, M. (1978). Highly repeated sequences of simian DNA. Science *200,* 394–402.

Sandeen, G., Wood, W. I., and Felsenfeld, G. (1980). The interaction of proteins HMG14 and 17 with nucleosomes. Nucl. Acids Res. *8,* 3757–3778.

Schmid, C. W., and Jelinek, W. R. (1982). The Alu family of dispersed repetitive sequences. Science *216,* 1065–1070.

Schröter, H., and Bode, J. (1982). The nucleosome binding sites for large and small HMG proteins. Eur. J. Biochem. *127,* 429–436.

Simpson, R. T., and Stafford, D. W. (1983). Structural features of a phased nucleosome core particle. Proc. Nat. Acad. Sci. USA *80,* 51–55.

Singer, M. (1982). Highly repeated sequences in mammalian genomes. Int. Rev. Cytol. *76,* 67–112.

Stein, A., and Townsend, T. (1983). HMG14/17 binding affinities and DNase I sensitivities of nucleoprotein particles. Nucl. Acids Res. *11,* 6803–6819.

Strauss, F., and Prunell, A. (1982). Nucleosome spacing in rat liver chromatin: a study with exonuclease III. Nucl. Acids Res. *10,* 2275–2302.

Streeck, R. E., Moritz, K. B., and Beer, K. (1983). Chromatin elimination in Ascaris suum: the sequence of the eliminated satellite DNA. Nucl. Acids Res. *10,* 3495–3502.

Swerdlow, P., and Varshavsky, A. (1983). Affinity of HMG17 for a mononucleosome is not influenced by the presence of ubiquitin-H2A semihistone but strongly depends on DNA fragment size. Nucl. Acids Res. *11,* 387–401.

Thomas, J. O., and Kornberg, R. D. (1975). An octamer of histones in chromatin and free in solution. Proc. Nat. Acad. Sci. USA *72,* 2626–2630.

Trifonov, E., and Sussman, J. L. (1980). The pitch of chromatin DNA is reflected in its nucleotide sequence. Proc. Nat. Acad. Sci. USA *77,* 3816–3820.

Varshavsky, A., Bakayev, V. V., and Georgiev, G. P. (1976). Heterogeneity of chromatin subunits and location of histone H1. Nucl. Acids Res. *3,* 477–492.

Varshavsky, A., Sundin, O., and Bohn, M. (1979). A stretch of "late" SV40 viral DNA about 400 bp long which includes the origin of replication is specifically exposed in SV40 minichromosomes. Cell *16,* 453–466.

Varshavsky, A., Levinger, L., Sundin, O., Barsoum, J., Ozkaynak, O., Swerdlow, P. and Finley, D. (1983). Cellular and SV40 chromatin: replication, segregation, ubiquitination, nuclease-hypersensitive sites, HMG-containing nucleosomes, and heterochromatin-specific protein. Cold Spring Harbor Symp. Quant. Biol. *47,* 511–528.

von Hippel, P. H. (1979). On the molecular bases of the specificity of interaction of transcriptional proteins with genome DNA. In Biological Regulation and Development, *1,* R. F. Goldberger, ed. (New York: Plenum Press), pp. 279–347.

Walker, J. M. (1982). Primary structures of HMG proteins. In The HMG Chromosomal Proteins, E. W. Johns, ed. (London: Academic Press), pp. 69–88.

Wasylyk, B., Oudet, P., and Chambon, P. (1979). Preferential in vitro assembly of nucleosome cores on some AT-rich regions of SV40 DNA. Nucl. Acids Res. *7,* 705–713.

Wittig, S., and Wittig, B. (1982). Function of tRNA gene promoter depends on nucleosome position. Nature *297,* 31–38.

Worcel, A., Strogatz, S., and Riley, D. (1981). Structure of chromatin and the linking number of DNA. Proc. Nat. Acad. Sci. USA *78,* 1461–1465.

Worcel, A. Gargiulo, G., Jessee, B., Udvardy, A., Louis, C., and Schedl, P. (1983). Chromatin fine structure of the histone gene complex of Drosophila melanogaster. Nucl. Acids Res. *11,* 421–432.

Wu, K., Strauss, F., and Varshavsky, A. (1983). Nucleosome arrangement in green monkey α-satellite chromatin. J. Mol. Biol. *170,* 93–117.

Young, D., and Carroll, D. (1983). Regular arrangement of nucleosomes on 5S RNA genes in Xenopus laevis. Mol. Cell. Biol. *3,* 720–730.

Zhang, X. Y., Fittler, F., and Hörz, W. (1983). Eight different highly specific nucleosome phases on α-satellite DNA in African green monkey. Nucl. Acids Res. *11,* 4287–4305.

Zhurkin, V. B. (1983). Specific alignment of nucleosomes on DNA correlates with periodic distribution of purine–pyrimidine and pyrimidine–purine dimers. FEBS Lett. *158,* 293–297.