Two target sites for protein binding in the promoter region of a cell cycle regulated human H1 histone gene

Andre J. van Wijnen, Kenneth L. Wright, Robert F. Massung, Martijn Gerretsen, Janet L. Stein and Gary S. Stein

Department of Cell Biology, University of Massachusetts Medical Center, Worcester, MA 01655, USA

Received August 28, 1987; Revised and Accepted December 14, 1987

ABSTRACT

The 5' region of a cell cycle regulated human H1 histone gene appears to contain at least six promoter DNA elements that are shared with some, but not all human core (H2A, H2B, H3 and H4) histone genes. We show that two of these elements represent separate binding sites for two distinct, partially purified factors. The first promoter domain contains A/T rich repeats and is involved in the binding of H1NF-A, a nuclear factor previously found to bind to A/T rich direct repeats in the promoters of human H4 and H3 histone genes. The second domain, containing the general promoter element 5' dACCAAT, acts as a binding site for a two component mosaic factor we have designated H1NF-B. These data suggest that coordinate transcriptional regulation of human H1 and core histone genes may involve two classes of trans-acting factors: those specific for histone gene promoters and those that act on a broad spectrum of human gene promoters.

INTRODUCTION

Human H1 and core histone genes comprise a set of genes that reside in multiple, polymorphically arranged clusters (1-4) located on at least three different chromosomes (5-6). Coordinate expression of cell cycle regulated human H1 (7-8) and core histone (7-12) genes is temporally and functionally coupled to the process of DNA replication during S-phase, and histone mRNA abundance is regulated at both transcriptional and post transcriptional levels (reviewed in 13-14). Transcriptional regulation involves a 3 to 5 fold transient enhancement of the rate of histone gene transcription at the G1/S-phase boundary over a basal rate that continues throughout the cell cycle (10, 12).

The mechanisms governing transcriptional control of human histone gene expression, as exemplified by cell cycle dependent human H4 and H3 histone genes, have been probed both *in vivo* and *in vitro* by a variety of techniques (15-22). This work revealed that the S-phase specific increase in H4 histone gene transcription rates is accompanied by significant, reversible alterations in chromatin organization of the H4 histone gene promoter (15-16).

Transcription studies performed *in vivo* (17-19) and *in vitro* (20) have delineated proximal and distal promoter elements implicated in the regulation of human H4 histone gene transcription. The proximal promoter elements of this gene have been correlated with two DNA/protein interaction sites that are detected both *in vivo* (21) and *in vitro* (22) and at least three distinct factors involved in these interactions have been partially purified (22 and unpublished data). *In vitro* studies performed by other workers have yielded interesting data on the transcriptional regulation of two other human core histone genes (23-25), but equivalent information is not available for human H1 histone genes.

Coordinate expression of multiple H1 and core histone genes in tight conjunction with DNA synthesis is a phenomenon that is not well understood, although it is conceivable that a common mechanism underlies the regulation of these genes (7-8). Transcriptional regulation in general is mediated, at least in part, by DNA binding proteins that bind to distinct promoter elements (26). Thus, control at this level may require regulatory factors that are involved in coordinating histone gene transcription during the cell cycle. Although many promoter elements (18-25) and various nuclear factors have been characterized so far (22-24), none of the factors has been studied in sufficient detail to unequivocally establish a role for these factors (or their cognate cis-acting elements) for coordination and S-phase specific enhancement of transcription as observed *in vivo*.

In this work we have established the functionality of the 5' flanking region of a cell cycle dependent human H1 histone gene designated FNC16 (1) by S1 nuclease protection analysis and determined the DNA sequence of a 0.5 kB fragment preceding the gene. By using a variety of sequence specific DNA/protein binding assays we have characterized two target sites for protein binding in the H1 histone gene promoter and partially purified the factors involved. Together, these results enable us to make a comparative analysis of factor binding sites in both H1 and core human histone gene promoters and to formulate a working model for the coordinate transcriptional regulation of human H1 and core histone genes.

MATERIALS AND METHODS

Vectors and DNA sequence analysis

The human H1 histone gene was derived from a genomic histone gene cluster cloned into lambda Charon 4A (1) and subcloned into pBR322 to generate the plasmids pFNC16 and pFNC16A (3). Restriction fragments spanning the 5' region

were subcloned into M13mp19 and subjected to Sanger dideoxy sequence analysis (27). The most proximal part of the promoter DNA sequence was confirmed by Maxam & Gilbert reactions (28). Plasmid pOX001 was prepared by subcloning an upstream SmaI/PvuII fragment of pFNC16A into Bluescript KS(M13+) (Stratagene).

Nuclear extracts and fractionation

Preparation of nuclear extracts (NE) and fractionation of these by DEAE-Sephacel and phosphocellulose column chromatography was performed as described before (22), with the following exceptions. Freshly prepared nuclei from 8×10^9 exponentially growing HeLa S3 cells were washed once with 50 ml sucrose buffer (50 mM Hepes-NaOH, pH 7.5; 10 % sucrose) and extracted with the same buffer containing 400 mM KCl. Nuclear extracts were diluted to a concentration of 200 mM KCl (50 ml, 140 mg protein) and passed over a DEAE-Sephacel column equilibrated with buffer A (22) containing 200 mM KCl, and the flow through was collected (D0-200 fraction; 85 ml, 122 mg protein). The D0-200 fraction was further fractionated by passage over a phosphocellulose column equilibrated with buffer A containing 200 mM KCl and the flow through designated as the P0-200 fraction (100 ml, 58 mg protein). The column was subsequently washed with buffer A containing 350 mM KCl (P200-350 fraction; 23 ml, 17 mg) and then with buffer A containing 1000 mM KCl (P350-1000 fraction; 10 ml, 14 mg).

S1 nuclease protection analysis

Isolation of total cellular RNA from HeLa S3 cells synchronized by double thymidine block (10) and S1 nuclease protection analysis (18) were performed as described before. S-phase RNA and RNA outside of S-phase were isolated from cells at 5 hr and 11 hr respectively after release from the block. After treatment of cells for 1 hr with 1mM hydroxyurea, RNA was isolated from cells harvested 4 hr after release. The DNA probe used in these experiments was an end labelled PvuII/StuI fragment (size 731 nt) which yields a protected fragment of 260 nt (Fig. 1). The site of mRNA initiation was determined using an end labelled PvuII/BclI fragment (size 680 nt) which yields a protected fragment of 199 nt (unpublished data).

Gel retardation assays and stairway assays

Electrophoretic mobility shift assays (gel retardation assays) and reciprocal restriction site deletion analysis (stairway assay) were performed as described (22), with the amount and type of competitor DNA indicated in the figure legends. The DNA probes used in this study to detect DNA/protein interactions in vitro were restriction fragments derived from plasmid pOX001 that were single end labeled at a SmaI-site (nt -78) or at a RsaI-site (nt-223) (Fig. 1).

Fig. 1: Restriction map of the human genomic histone gene cluster lambda 49, showing the direction of H1 mRNA transcription (arrow) (top). Depicted below are restriction sub-fragments from the EcoRI fragment designated FNC16 used for S1 analysis (middle) and DNA binding studies (bottom).

DNaseI protection, DMS protection and methylation interference

DNaseI protection experiments were performed by making a 25-fold scaled up binding reaction as above, incubating the sample with 5u DNaseI (Boehringer) for 1 minute at room temperature and stopping the reaction by adding EDTA to a final concentration of 25 mM. Preparative separation of unbound probe and bound DNA/protein complexes by electrophoresis was followed by excision of the bands from the gel and elution by diffusion (27). After processing of the eluate the samples were analyzed on a denaturing 8% polyacrylamide gel containing 7M urea. DMS protection studies were done in a similar fashion with the exception that pulse incubation occurred by adding 2.5 μ l undiluted dimethylsulphate (DMS) and that no EDTA was added. After elution samples were treated with piperidine and guanine specific cleavage products were analyzed on gels (28). Methylation interference (29) was performed by incubation of the probe with DMS prior to the binding reaction. After removal of the DMS, the partially methylated probe was incubated with protein fractions and then the bound and unbound DNA fragments were separated by preparative gel electrophoresis. Subsequently, the samples were treated in the same way as for DMS-protection studies.

RESULTS

The human H1 histone gene FNC16 is functional and expressed in vivo

The extent to which expression of a cloned human H1 histone gene designated FNC16 (Fig. 1) is coupled to cell cycle progression was investigated by S1 nuclease protection analysis. It had been shown previously by Northern blot analysis with the FNC16 H1 histone gene as a hybridization probe, that human H1 histone gene expression is cell cycle dependent (10). However, cross

Fig. 2: S1 analysis of pFNC16A (100 ng) using mRNA (25 µg) isolated from synchronized HeLa S3 cells. mRNA was isolated at 5 hr (ln 1-2) and 10 hr (ln 3-4) after release from a double thymidine block. In 5-6: mRNA samples (5 hr) from cells treated with the DNA synthesis inhibitor hydroxy urea.

hybridization with other H1 histone mRNAs made it uncertain that the FNC16 H1 histone gene itself is expressed in a cell cycle dependent manner. In addition, functional copies of human histone genes are clustered with highly homologous, non expressed pseudo-genes (30), raising the possibility that the H1 histone gene FNC16 could be non-functional.

Therefore, we investigated the expression of the H1 histone gene FNC16 by S1 nuclease protection analysis of the RNAs isolated from synchronized HeLa S3 cells (Fig. 2). The probe used for these experiments was a 731 nt PvuII/StuI fragment of pFNC16, which includes 5' flanking sequences and the initial coding sequences of the H1 histone gene (Fig. 1) RNA isolated at the peak of S-phase (Fig. 2; ln 1-2) protected a 260 nt fragment of this probe from S1 nuclease digestion. The protected fragment is distinguishable from other S1 nuclease degradation products that can arise from hybridization of pFNC16 with homologous mRNA species. Very limited S1 protection occurs with RNA isolated from cells outside S-phase (Fig. 2; ln 3-4) or by S-phase RNA from cells treated with the DNA synthesis inhibitor hydroxyurea (Fig. 2; ln 5-6). Combined with previous work (10,11), this result establishes that the H1 histone gene FNC16 is a bona fide cell cycle regulated gene with expression tightly coupled to DNA replication.

Fig. 3: DNA sequence of the promoter of the human H1 histone gene FNC16. Overlined sequences are discussed in the text. Indicated are the restriction sites (SmaI and RsaI) used for end labeling of probes for DNA binding assays. Numbering is from the ATG-startcodon. Shown underneath is the sequencing strategy.

The H1 histone promoter contains both general and histone specific consensus transcriptional elements

The 5'flanking region of the human H1 histone gene FNC16 (1) has been analyzed by dideoxy sequencing (Fig. 3) to establish the presence of general RNA pol II transcriptional elements (26) and to determine the representation of histone specific consensus elements (31). In the immediate upstream region at least three general consensus cis-acting elements are present: (A) a highly symmetrical TATA-box (14 bp palindrome, nt -92 to -79), (B) a perfect copy of the PuCCAAT-box (nt -112 to -107) and (C) a GGGGG-box (nt -134 to -125); in addition, an inverted PuCCAAT-like element (nt -119 to -124) is contained within a region that has striking sequence symmetry (nt -139 and nt -106; two staggered palindromes) (see Results).

Fig. 4: Gel retardation assay using a crude nuclear extract in the presence of *E. coli* DNA (2 μg) (part A) or poly I/C DNA (2 μg) (part B). In 0–9 (part A and B): respectively, 0, 0.3, 0.6, 0.8, 1.1, 1.4, 1.7, 3.4, 4.5 and 5.6 μg protein added. Indicated by arrows are complex A, B, C and D.

Located more upstream from the above consensus elements are two perfect copies of a putative H1 specific consensus element (AAACACA-box; nt -173 to -167 and nt -282 to -276) (31). In addition, three A/T-rich repeats (between nt -198 to -185) can be observed; similar elements in the promoter of cell cycle dependent human H4 and H3 histone genes have been implicated in the binding of a factor designated HiNF-A (22). Another remarkable feature is an imperfect 20 bp repeat containing a DraI-site, starting at positions nt -323 and nt -430 respectively. From all of the above observations we inferred that the region between nt -80 to -200 is analogous to the upstream region between nt -50 to -150 of the cell cycle regulated H4 histone gene F0108 (21) that contains two major *in vivo* DNA/protein interaction sites.

Four DNA/protein complexes can be formed in the H1 histone promoter region *in vitro*

The gel retardation assay has been used to study DNA/protein interactions in the human H1 histone promoter *in vitro* in order to detect factors that are analogous or identical to DNA binding transcription factors observed for core H4 and H3 histone gene promoters. The probes used cover the region nt -233 to -78; this region contains at least one copy of all the consensus DNA elements mentioned above. In an initial experiment an end labelled SmaI/AluI restriction

Fig. 5A: Gel retardation assay using protein fractions in the presence of poly I/C DNA (2 μ g). In A1-A2: 1.4 and 2.2 μ g NE protein added; In B1-B2: 0.6 and 2.1 μ g D0-200 protein; In C1-C6: 0.2, 0.9, 1.7, 2.6, 3.5 and 4.4 μ g P0-200 protein; In D1-D6: 0.3, 1.1, 2.2, 3.3, 4.4 and 5.5 μ g P200-350 protein; In E1-E2: 0.7 and 1.4 μ g P350-1000 protein. **Fig. 5B:** Same as Fig. 5A, in the presence of E. coli DNA (2 μ g). In A1-A6: 0.6, 1.4, 2.1, 2.8, 4.2 and 5.6 μ g D0-200 protein added; In B1-B6: 0.2, 0.6, 0.9, 1.2, 1.8 and 2.4 μ g P0-200 protein; In C1-C6: 0.6, 0.9, 1.2, 1.5, 2.2 and 2.9 μ g P200-350 protein; In D1-D6: 0.3, 0.4, 0.5, 0.7, 1.1 and 1.4 μ g P350-1000 protein.

fragment (spanning nt -233 to -78) was incubated with increasing amounts of crude HeLa nuclear extract protein in the presence of either poly I/C DNA or E. coli DNA. Upon electrophoresis of the mixtures containing E. coli DNA in a low ionic strength, native polyacrylamide gel, two prominent retarded DNA fragments (designated complex A and complex B) were observed (Fig. 4; ln A1-A9). However, if poly I/C DNA was used as competitor DNA, only the upper retarded fragment (complex B) and two other retarded fragments (designated complex C and complex D) were detectable (Fig. 4; ln B1-B9). Apart from the above mentioned DNA/protein complexes we observed an additional complex slightly above complex B (not indicated). Similar satellite bands are also associated with major H4 and H3 histone promoter/protein complexes (unpublished observations). In these cases, formation of such DNA/protein complexes requires the same DNA sequences and the factors involved elute in the same fractions; this suggest that satellite bands are formed by closely related factors. The factor involved in the satellite band of complex B meets similar criteria (see Fig. 4 and 5; not indicated).

Complex B displayed an intensity not previously observed for any of the probes used in our gel retardation assays (22 and unpublished data) suggesting that the factor bound to the probe is relatively abundant and/or has a high affinity for the H1 histone gene promoter. Complexes C and D could represent separate nuclear factors and/or oligomerization states of factors present in complex B or complex A.

Complex A showed similarities in relative migration rate and intensity with a complex formed by a factor designated HiNF-A (see previous section) that binds to H3 and H4 histone gene promoters. In addition, as is the case here, binding of HiNF-A to these promoters was not detected when poly I/C DNA was included in the binding reaction, suggesting that the factor interacting with the human H1 histone gene promoter is similar or identical to HiNF-A.

Two distinct factors interact with the H1 histone promoter

Nuclear extracts have been fractionated by liquid column chromatography on DEAE-Sephacel and phosphocellulose in order to separate the DNA binding activities that associate with human histone gene promoters and to obtain qualitative data on the physical properties of these factors. We have obtained a DEAE-Sephacel flow through fraction (D0-200; see Materials and Methods) and have separated the D0-200 fraction into three phosphocellulose fractions (P0-200, P200-350 and P350-1000 fractions, respectively).

The gel retardation assay was used to monitor the elution behavior of the

factors involved in formation of complexes A, B, C and D. Incubation of the SmaI/AluI HI-probe with the P0-200 fraction in the presence of poly I/C DNA resulted in the formation of complexes B and C (Fig. 5A; ln C1-C6). Residual formation of complex B, but not of complex C, was observed with the P200-350 fraction (Fig. 5A; ln D1-D6), but no detectable formation of either complex was obtained with the P350-1000 fraction (Fig. 5A; ln E1-E2). These data indicate that the factor involved in complex B formation (designated HiNF-B) elutes primarily in the P0-200 fraction and that the factor(s) involved in complex C co-elutes with HiNF-B in the P0-200 fraction. Formation of complex D was observed for both the P0-200 and P200-350 fractions (Fig. 5A; ln C1-C6 and D1-D6) showing that the factor(s) involved only partially co-elutes with HiNF-B. This suggests that complex D is unrelated to complexes B and C. In addition to the above complexes, we observed a novel retarded band (designated complex B') with the P0-200 or P200-350 fractions (see below).

Complex A formation was analyzed using the same protein fractions by including poly I/C DNA in stead of *E. coli* DNA in the binding reaction. Formation of complex A was observed with the P350-1000 fraction (Fig. 5B; ln D1-D6), but not with the other phosphocellulose fractions (Fig. 5B; ln B1-B6 and C1-C6). The detection of complex A in the P350-1000 fraction in the presence of *E. coli* DNA, but not poly I/C DNA (see Fig. 5A; ln E1-2), again suggests that the factor involved in formation of complex A is similar or identical to HiNF-A. Because this factor fits the current operational definition for HiNF-A we have given it the same designation, although it could be that HiNF-A is a collection of closely related factors.

In summary, we have detected at least two distinct factors (HiNF-A and HiNF-B) involved in the formation of two prominent DNA/protein complexes (complex A and complex B). Moreover, our results indicate another DNA/protein complex that may be related to factor HiNF-B (complex C), but probably not HiNF-A, and a complex that may be unrelated to either factor HiNF-A or factor HiNF-B (complex D).

Factor HiNF-B consists of two components, B1 and B2

Incubation of the P0-200 fraction in the presence of *E. coli* DNA resulted in a drastic decrease in the formation of complex B as compared to binding reactions containing poly I/C DNA (compare Fig. 5B, ln B1-B6 with Fig. 5A, ln C1-C6); in addition, a minor DNA/protein complex (designated B') was much more prominent when *E. coli* DNA was substituted for poly I/C DNA as competitor in the binding reaction. We have obtained evidence that suggests that complex B' is an intermediate in the formation of complex B and that formation of complex

Fig. 6: Gel retardation assay showing reconstitution of HiNF-B activity. In 1: 2.4 μ g P0-200 protein added; In 2: 2.8 μ g P200-350 protein; In 3: combination of 1.2 μ g P0-200 protein and 1.4 μ g P200-350 protein; *E. coli* DNA (2 μ g) as competitor.

B requires the combination of two components B1 and B2 (unpublished data; 32).

When *E. coli* DNA was used in the assay, HiNF-B DNA binding activity could be reconstituted by combining the P0-200 fraction with the P200-350 fraction (Fig. 6; very light exposure allowing densitometric scanning). For instance, the combination of 1.2 μ g P0-200 protein and 1.4 μ g P200-350 protein (2.6 μ g total protein) resulted in a 4 to 5 fold enhancement of complex B formation as compared to 2.4 μ g P0-200 protein alone; also, complex B was virtually undetectable with 2.8 μ g P200-350 protein alone. This finding supports the notion that HiNF-B is a mosaic factor consisting of at least two components.

Factor HiNF-A binds to a distal promoter segment containing AT-rich repeats

Reciprocal restriction site deletion analysis in combination with the gel retardation assay ("stairway assay") was used to fine map the segment in the human H1 histone promoter involved in formation of complex A (Fig. 7A and 7B). Restriction fragments labelled either at a *Sma*I site (nt -76) or at a *Rsa*I site (nt -213) were prepared and shortened by various restriction enzymes as indicated in the figure legends; these probes were used in parallel in gel retardation assays. Incubation of shortened probes labelled at the *Sma*I site with P350-1000 protein resulted in formation of complex A when sequences up to an *Rsa*I site were deleted (Fig. 7; In B1-B6), but not when sequences upstream of an *Hinf*I site were omitted, showing that sequences between the *Rsa*I site and the *Hinf*I site (nt -213 to -160) are required for complex A formation.

Fig. 7A and 7B: Stairway assay of probes end labelled (indicated by black star) at the RsaI site (panel A) and SmaI site (panel B). In A1-A3 and B1-B3: In 1-3, no protein added; panel A: RsaI probes digested with MspI (nt -93), HaeIII (nt -139) and HinfI (nt -150); panel B: SmaI probes digested with AluI (nt -233), RsaI (nt -213) and HinfI (nt -150). Panel A and B: In 4-6, 1 μ g P350-1000 protein added; E. coli DNA as competitor. **Fig. 7C and 7D:** See legend 7A and 7B (probes as indicated and poly I/C DNA as competitor). In C1-C4 and D1-D3: 2.1 μ g D0-200 protein added; In C5-C6: 1.2 μ g P0-200 protein.

The reciprocal experiment was carried out in which the various probes were labelled at the RsaI site (Fig. 7; In A1-A6). Binding reactions performed with these probes in the presence of P350-1000 protein revealed that sequences

downstream of a HinfI site (nt -160) are dispensable for factor binding and that the RsaI/HinfI fragment (54 bp) by itself is capable of binding the factor. Thus, we have identified a DNA segment (nt -213 to -160) that is both required and sufficient for binding of HiNF-A. Inspection of the DNA sequence of this fragment shows that it contains two direct and one inverted repeat of an ATTT-element. Similar repeat elements have been shown to bind HiNF-A in human H3 and H4 histone promoters (22). Therefore, the assignment that complex A in the H1 histone promoter is formed by HiNF-A or a related factor is consistent with these findings.

Factor HiNF-B binds to the most proximal promoter sequences

Stairway assays were also used to localize the binding domain for HiNF-B in the H1 histone promoter. Probes labelled at the SmaI site were incubated with a fixed amount of protein and analyzed on gel (Fig. 7C and D). Deletion of sequences upstream of a HaeIII site (nt -138) had no effect on formation of complex B and a HaeIII/SmaI fragment by itself was sufficient for binding of HiNF-B (Fig. 7; ln C1-C4, D0-200 fraction, and ln C5-C8, P0-200 fraction). Hence, we have delineated a 53 bp fragment (nt -138 to -76) that is sufficient for binding of HiNF-B.

Reciprocal probes labelled at the RsaI sites were also analyzed in the assay. In accordance with the results presented above, only a RsaI/SmaI fragment yields substantial complex B formation, although residual binding is still observed with a RsaI/HinfI fragment (Fig. 7; ln D1-D3). The latter suggests the presence of an additional, low affinity binding site for HiNF-B in the 3' portion of the probe between nt -213 and -160. The results obtained for complex C (Fig. 7; ln C1-C4 and D1-D3) do not seem to be compatible with a single binding domain for a unique factor. This complex is detected only when a substantial amount of complex B is formed, further suggesting that complex B and C are related.

Factor HiNF-B is a PuCCAAT-box binding protein

The DNA segment (nt -138 to -76) involved in binding of HiNF-B contains three consensus transcriptional elements (see previous section) necessitating a more detailed analysis of the exact HiNF-B binding site. This was accomplished by DNaseI footprint analysis in which H1 probes end-labelled at the top (sense/+) or bottom (anti sense/-) strand were incubated with D0-200 protein fraction and mixtures pulse digested with DNaseI (Fig. 8A). Separation of complex B from unbound DNA and unstable DNA/protein complexes (background smearing on gels) was accomplished by native polyacrylamide gel electrophoresis. The radioactive DNA in the gel segments corresponding to

Figure 8: DNaseI footprinting (panel A), DMS fingerprinting and DMS interference (panel B) of the HINF-B binding site. In A1-A2 [(-) strand], and in A5-A6 [(+) strand], show M & G sequencing reactions as indicated; (-) strand: DNA bound by HINF-B (ln 3: B) and unbound DNA (ln 4: U); (+) strand: DNA bound by HINF-B (ln 7: B) and unbound DNA (ln 8: U). The maximum DNaseI footprint boundaries are indicated by brackets, the PuCCAAT-box is shown for reference. Panel B: ln 1 and 5, duplicate G-reactions of (-) strand; ln 2,

unbound DNA (U) and ln 3-4, duplicate HiNF-B bound DNA (B). DMS protected bases are indicated by open circles (size is measure for protection), strong DMS enhancement is shown by a black dot. Dashes indicate G's at nt -118 to -90 shown for reference. Panel B: ln 6, G-reaction of (-) strand; ln 7, unbound DNA (U) and ln 8, HiNF-B bound DNA (B). Indicated by arrows: nucleotides enriched in unbound DNA fractions but depleted in HiNF-B bound DNA fractions.

complex B (bound DNA) or unbound DNA was analyzed by denaturing gel electrophoresis. Using this approach, the minimal boundaries for protection by HiNF-B against DNaseI phosphodiester cleavage were established at nt -120 and -92 on the top (+) strand and between nt -121 and -90 on the bottom strand (Fig. 8A), although differences were detected in the extent to which HiNF-B prevents accessibility of DNaseI to particular residues within the HiNF-B binding site. The DNA sequence of this DNaseI footprint contains a perfect PuCCAAT motif, but lacks the other consensus elements. Therefore, we conclude that HiNF-B is a PuCCAAT-box binding protein (see Discussion). In addition, the HiNF-B footprint extends further towards the 3' side of the PuCCAAT-box and terminates at the first nucleotides of a palindromic TATA-element which may have important biological implications (see next section).

No protection was observed for an inverted PuCCAAT-box located at nt -119 to -124 despite the fact that this element is present at a distance from the protected PuCCAAT-box that would not prohibit dimerization of two single PuCCAAT-box factors (33, 34). Taking into account that HiNF-B is a two component factor it is for this reason unlikely that HiNF-B is a dimer of a single CCAAT-box factor.

Contact point analysis of the HiNF-B binding site

The interaction of HiNF-B with its binding site was studied at single nucleotide resolution by DMS fingerprinting in order to define points of close contact between HiNF-B and the guanine bases within its DNA recognition sequence. This approach, which measures the inaccessibility of purine-residues to methylation by dimethylsulphate during the binding reaction (Fig. 8B), was necessary in order to confirm the presence of PuCCAAT-box specific DNA/protein contacts and to uncover possible additional contacts within the footprint. To this end, the bottom (-) strand, containing the majority of G-residues within the footprint, was end labeled and this double stranded H1 probe was pre-incubated in the presence of the D0-200 fraction and then treated with DMS in a manner similar to that used for DNaseI footprinting experiments.

DMS treated samples derived from the DNA fraction representing complex B (bound) showed reproducible protection of G-residues nt -111 and -110, the complementary nucleotides of the PuCCAAT-box, and detectable hypomethylation of G-nucleotide -99, localized in the 3' side of the DNaseI footprint. Occasionally, we observed faint protection of G-104 by comparing the intensity ratio of this G with the unprotected residue G-106. We noted a very strong enhancement of methylation of G-residue nt -113 suggesting that binding of HiNF-B makes this nucleotide highly accessible to DMS attack. In conclusion,

the DMS protection pattern is compatible with the assignment of HiNF-B as a PuCCAAT-box binding protein, but the presence of additional contacts in the 3' side of the DNaseI footprint supports the idea that a second DNA binding activity is present, that binds in conjunction with the PuCCAAT-box factor.

The binding site of HiNF-B was studied in further detail by DMS methylation interference experiments. These studies were carried out to assess whether the methylation of certain G-residues by DMS treatment prior to the binding reaction, interfered with binding of HiNF-B (Fig. 8B). Again, the bottom (-) strand was end-labelled, the DNA fragment complexed to HiNF-B and the radioactive DNA fractions (bound and unbound) were isolated in the same way as in the above experiments. The results obtained showed clearly that methylation of G-residues nt -111 and -110, but no other G-residues, abolished binding because the corresponding bands were missing in the G-specific piperidine cleavage pattern of the bound fraction and were significantly enriched in the uncomplexed DNA fraction. Because DMS methylates G-residues at the major groove of the DNA helix, the data derived from the DMS fingerprinting and DMS methylation interference experiments demonstrate that HiNF-B binds in close proximity to the major groove of the PuCCAAT-box (Fig. 9 summarizes the above data). However, no interference occurred at any of the other G-residues within the bottom strand implying that the methylation protection and methylation enhancement of the G-residues peripheral to the PuCCAAT-box are caused by funneling and shielding effects intrinsic to the physical surface of the bound factor.

Vertebrate histone gene promoter DNA sequences (31) were searched for similarity with the DNA sequence contained within the HiNF-B footprint. This analysis showed that the HiNF-B binding site contains the Breathnach/Chambon CCAAT-consensus sequence (35), but in histone gene promoters the homology extends further; the consensus CCAAT-element for histone gene promoters is 5'dPyPyPuPuCCAAT(C/G)APuPuPu.

Comparative analysis of human H1 and core H3 and H4 histone gene promoter DNA elements and protein binding sites

The promoter regions of three cell cycle dependent human histone genes (H1, H3 and H4) have been investigated for DNA/protein interactions and, in total, we have detected at least six DNA/protein interaction sites in vitro (22 and unpublished data) some of which have been shown to overlap *in vivo* protein binding sites (21); a summary of these findings is shown in Figure 10. Factor HiNF-A appears to have binding sites in all three histone gene promoters examined and protection of its binding site in the H4 histone gene promoter

Figure 9: Binding site of HiNF-B determined on (+) and (-) strand by DNaseI footprinting (black lines; dotted line: no DNaseI cuts) and DMS protection/-interference on (-) strand (symbols as in Fig. 8).

from DNaseI and DMS *in vivo* has been established by genomic footprinting and fingerprinting (21). The conservation of HiNF-A binding sites in histone gene promoters clearly suggests a specific role in the regulation of histone gene transcription.

Factor HiNF-B appears to bind to a region immediately upstream of the TATA-box in the H1 histone gene promoter, which is analogous to the *in vivo* binding domain Site II in the H4 histone gene promoter (21). The distal part (5' side) of Site II immediately upstream of the TATA-box is absolutely required for efficient accurate transcription initiation of this H4 histone gene *in vivo* (18). Hence, by analogy with the *in vivo* protein binding site in the H4 histone promoter it is possible that the HiNF-B binding site in the H1 histone promoter serves a similar role.

DISCUSSION

In this work two distinct nuclear factors have been shown to bind independently to a proximal promoter element (HiNF-B) and a more distal element (HiNF-A) in the immediate upstream region of a cell cycle regulated human H1

Figure 10: Established factor binding sites in the proximal promoters of human H1 (FNC16), H3 (ST519) and H4 (FO108) histone genes. Indicated are the presence of TATA-, CCAAT- and GC-homologies (26) and histone specific AAACACA- and GGTCG-elements (thin lines) (31). *In vitro* factor binding sites (22, 42, this work) are overlined and *in vivo* binding sites (I and II) for the H4 histone promoter are indicated by thick lines underneath the sequence (21).

histone gene. HiNF-A is a protein that causes slight retardation of DNA fragments containing its binding site in gel retardation assays, suggesting that it is a rather small protein. The factor is present in the high salt phosphocellulose fraction of a nuclear extract which represents a small part of total nuclear protein. The binding of HiNF-A is undetectable in the presence of low amounts of poly I/C DNA, but is relatively resistant to competition by *E. coli* DNA in gel retardation assays suggesting that poly I/C DNA has certain features that resemble its binding site. The factor has a sequence specific DNA binding activity as has been demonstrated, for instance, by stairway assays and DNaseI footprinting (22).

Thus far, the only documented protein that resembles HiNF-A in some aspects is mouse alpha protein (36); this protein is comparably small, elutes in a similar high salt phosphocellulose fraction, is competed out specifically with poly I/C DNA but not with other synthetic DNA substrates and recognizes a series of apparently random A/T rich sequences. The latter property is the only feature that does not seem to be shared between alpha protein and HiNF-A; rather, HiNF-A appears to bind to small repeated ATTT-elements that are present at comparable locations in the three histone gene promoters we have investigated to date. So far, we have not been able to detect a direct role for the protein in transcription *in vitro* (K. Wright, unpublished data), although the conserved location of its site in both human H1 and core histone gene promoters and detection of its protected site *in vivo* (21) suggest an important biological role. The similarity of HiNF-A with mouse alpha protein may extend further; for instance, the latter protein is implicated in nucleosome positioning. In viewing the cell cycle dependent changes in chromatin structure that occur in the histone gene promoter during S-phase (15, 16) this leaves open an attractive possibility for a function of HiNF-A that can not be addressed easily by *in vitro* studies.

HiNF-B has been shown in this work to bind to a well conserved PuCCAAT-motif in the human H1 histone gene FNC16 that is present in a number of other histone gene promoters (31) as well as a broad spectrum of vertebrate promoters (26). A number of PuCCAAT-box binding proteins (CBP) have been described to date among which are mouse alpha globin CBP (ag-CBP) (37), Herpes simplex virus thymidine kinase CBP (tk-CBP) (38), human NF-I (39), vertebrate NF-I/TGOCA-binding proteins (40) and human NF-I/CITF (41). The latter three factors are independent CBP isolates from several different laboratories characterized initially by their ability to stimulate Adenovirus DNA replication (39), but

later shown to be analogous or identical to the CCAAT transcription factor (CTF) (41).

Our HiNF-B/CBP seems different from tk-CBP in that our factor is heat sensitive (32; discussion in 41) and different from tk-CBP and CTF/NF-I in DNaseI footprint size: HiNF-B has a footprint that extends more to the 3' side of the PuCCAAT-element, a property shared with (heat sensitive) ag-CBP (37). Comparison of the contact points of the PuCCAAT binding site with HiNF-B, ag-CBP (37) or NF-I (34) shows that the same G-residues are protected against dimethylsulphate; although we have found additional protection and enhancement of G's in the H1 histone promoter, these G-residues are not present at corresponding positions in the other sites investigated (34, 37). Moreover, our methylation interference studies using HiNF-B have shown that analogous methylated G's in the HiNF-B and NF-I site (34) prohibit binding. We conclude that human HiNF-B, mouse alpha globin CBP and human NF-I have similar contact points with their cognate sequence, although their DNaseI footprints seem to show discrepancies.

Factors resembling HiNF-A and HiNF-B can be isolated from mouse cells (32), suggesting that mouse ag-CBP and human HiNF-B could be protein homologues. Comparison of the chromatographic behavior of ag-CBP and HiNF-B shows that the former elutes in the phosphocellulose flow through (100 mM KCl) (37). In contrast, HiNF-B exhibits extensive dissociation in a similar flow through fraction (200 mM KCl) into at least two components B1 and B2. A similar comparison between NF-I and HiNF-B reveals that NF-I elutes at 350 mM NaCl, with a minor peak of activity at 450 mM NaCl (39). Such data show that these factors are physically different. However, because HiNF-B is a two component factor the possibility must be considered that one of these components resembles other CBP's when separated from the second. In conclusion, HiNF-B appears to be a new member of a growing class of PuCCAAT-box DNA binding proteins.

The promoter regions of cell cycle regulated human H1 and core H4 and H3 histone genes cloned in our laboratory have been studied in detail both *in vivo* and *in vitro* (15-22, 32, 42 and this work). The picture that is emerging from these studies is that coordinate transcriptional control of histone gene expression may be exerted by a complex combination of three classes of trans-acting factors: (I) multiple, general transcription factors that bind to subsets of histone promoters in conjunction with both (II) histone specific factors, and possibly (III) less general, cell cycle stage specific factors. This notion is in agreement with the working model of Heintz and collaborators

(25) that is based on sequence homology between two (H2B and H4) core histone gene promoters. Regardless of the interrelationship between the various factors and their cognate promoter elements identified so far, the rigorous purification of histone promoter specific DNA binding activities by DNA affinity chromatography is a pre-requisite in addressing the validity of any transcriptional regulation model.

ACKNOWLEDGMENTS

This work was supported by grants from the National Institute of Health (GM 32010, GM 32381), the National Science Foundation (PCM 83-18177) and the March of Dimes Birth Defects Foundation (1-183). AJVW and MG are recipients of an exchange program grant from the University of Utrecht, Netherlands, under the supervision of Dr. Harry Voorma. We thank Wesley Faunce for performing initial experiments and Drs. Harry Nick, Peter van der Vliet and Erik de Vries for advice and materials. Special thanks to Urs Pauli and Susan Chrysogelos for critical discussions.

REFERENCES

1. Carozzi, N., Marashi, F., Plumb, M., Zimmerman, A., Coles, L.S., Wells, J.R.E., Stein, G. and Stein, J. (1984) *Science* **224**, 1115-1117.
2. Sierra, F., Lichtler, A., Marashi, F., Rickles, R., Van Dyke, T., Clark, S., Wells, J., and Stein, G., and Stein, J. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 1795-1799.
3. Zwollo, P., Stein, G.S. and Stein, J.L. (1984) *Biochem. Biophys. Res. Comm.* **124**, 988-993.
4. Heintz, N., Zernik, M. and Roeder, R.G. (1981) *Cell* **18**, 661-668.
5. Green, L., van Antwerpen, R., Stein, J., Stein, G., Triputti, P., Emanuel, B.S., Croce, C. (1984) *Science* **226**, 838-840.
6. Triputti, P., Emanuel, B.S., Croce, C.M., Green, L.G., Stein, G.S. and Stein, J.L. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 3185-3188.
7. Plumb, M., Marashi, F., Green, L., Zimmerman, A., Zimmerman, S., Stein, J. and Stein, G. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 434-438.
8. Baumbach, L., Marashi, F., Plumb, M., Stein, G. and Stein, J. (1984) *Biochem.* **23**, 1618-1625.
9. Lichtler, A.C., Sierra, F., Clark, S., Wells, J.R.E., Stein, G.S. and Stein, J.L. (1982) *Nature* **298**, 195-198.
10. Plumb, M., Stein, J. and Stein, G. (1983) *Nucl. Acids Res.* **11**, 2391-2410.
11. Plumb, M., Stein, J. and Stein, G. (1983) *Nucl. Acids Res.* **11**, 7927-7945.
12. Heintz, N., Sive, H.L. and Roeder, R. (1983) *Mol. Cell. Biol.* **3**, 539-550.
13. Stein, G.S., Stein, J.L. and Marzluff, W.F. (eds.) (1984) "Histone genes" John Wiley and Sons, New York.
14. Schumperli, D. (1986) *Cell* **45**, 471-472.
15. Chrysogelos, S., Riley, D.E., Stein, G. and Stein, J.L. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 7535-7539.
16. Moreno, M.L., Chrysogelos, S.A., Stein, G.S. and Stein, J.L. (1986) *Biochem.* **25**, 5364-5370.
17. Marashi, F., Helms, S., Shiels, A., Silverstein, S., Greenspan, D., Stein, J. and Stein, G. (1986) *Biochem. Cell Biol.* **64**, 277-289.
18. Kroeger, P., Stewart, C., Schaap, T., van Wijnen, A., Hirshman, J., Helms, S., Stein, G. and Stein, J. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 3982-3986.
19. Helms, S.R., van Wijnen, A.J., Kroeger, P., Shiels, A., Stewart, C.,

- Hirshman, J., Stein, J.L. and Stein, G.S. (1987) *J. Cell. Physiol.* **132**, 552-558.
20. Sierra, F., Stein, G. and Stein, J. (1983) *Nucl. Acids Res.* **11**, 7069-7086.
21. Pauli, U., Chrysogelos, S., Stein, G., Stein, J. and Nick, H. (1987) *Science* **236**, 1308-1311.
22. van Wijnen, A.J., Stein, J.L., and Stein, G.S. (1987) *Nucl. Acids Res.* **15**, 1679-1698.
23. Dailey, L., Hanly, S., Roeder, R.G. and Heintz, N. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 7241-7245.
24. Sive, H.L. and Roeder, R.G. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 6382-6386.
25. Sive, H.L., Heintz, N. and Roeder, R.G. (1986) *Mol. Cell. Biol.* **6**, 3329-3340.
26. Dynan, W. S. and Tjian, R. (1985) *Nature (London)* **316**, 774-778.
27. Messing, J. (1981) in: *Methods in Enzymology* **101** (part C) 20-78. Academic press, New York.
28. Maxam, A.M. and Gilbert, W. (1980) in: *Methods in enzymology* **65**, 499-560. Academic press, New York.
29. Siebenlist, U. and Gilbert, W. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 122-126.
30. Marashi, F., Prokopp, K., Stein, J. and Stein, G. (1984) *Proc. Natl. Acad. Sci.* **81**, 1936-1940.
31. Wells, D.E. (1986) *Nucl. Acids Res.* **14**, Supplement r119-149.
32. van Wijnen, A.J., Massung, R.F., Faunce, W. H. and Stein, J. L. and Stein, G. S. (1987) manuscript in preparation.
33. Gronostajski, R.M. (1987) *Nucl. Acids Res.* **15**, 5545-5559.
34. de Vries, E., van den Heuvel, S. J. L., van Driel, W. and van der Vliet, P. C. (1987) *EMBO J.* **6**, 161-168.
35. Breathnach, R. and Chambon (1981) *Annu. Rev. Biochem.* **50**, 349-383
36. Solomon, M.J., Strauss, F. and Varshavsky, A. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 1276-1280.
37. Cohen, R. B., Sheffery, M. and Kim, C. G. (1987) *Mol. Cell. Biol.* **6**, 821-832.
38. Graves, B. J., Johnson, P. F. and McKnight, S. L. (1986) *Cell* **44**, 565-576.
39. Nagata, K., Guggenheimer, R. A., Enomoto, T., Lichy, J. H. and Hurwitz, J. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 6438-6442.
40. Leegwater, P. A. J., van der Vliet, P. C., Rupp, R. A. W., Nowock, J. and Sippel, A.E. (1986) *EMBO J.* **5**, 381-386.
41. Jones, K. A., Kadonaga, J. T., Rosenfeld, P. J., Kelly, T. J. and Tjian, R. (1987) *Cell* **48**, 79-89.
42. van Wijnen, A. J., Wright, K. L., Kroeger, P., Stein, J. L. and Stein, G. S. (1987) manuscript submitted.