# Molecular Cloning of an Enhancer Binding Protein: Isolation by Screening of an Expression Library with a Recognition Site DNA

**Harinder Singh, Jonathan H. LeBowitz, Albert S. Baldwin, Jr., and Phillip A. Sharp**
Center for Cancer Research and Department of Biology
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

## Summary

**A novel strategy has been used to isolate a cDNA clone that encodes a DNA binding domain whose recognition properties overlap those of the mammalian transcription factors H2TF1 and NF-κB. These two factors are distinguished by their cell type distributions and their relative affinities for related sequence elements in the enhancers of the major histocompatibility complex (MHC) class I and immunoglobulin κ chain genes. The human cDNA clone was detected by screening a λ phage expression library with a binding site probe derived from the MHC enhancer. The phage encoded fusion protein binds specifically to both the MHC and κ gene enhancers. The cDNA hybridizes to a single copy gene that is expressed as a 10 kb mRNA in both B and non-B cells. The strategy used in this study may prove generally useful in the cloning and analysis of sequence-specific DNA binding proteins.**

## Introduction

To facilitate the isolation of genes specifying mammalian transcription factors, we have used a novel strategy designed to detect clones encoding sequence-specific DNA binding proteins. The strategy depends on the functional expression in E. coli of high levels of the DNA binding domain of a regulatory protein and a strong interaction between this domain and its recognition site. If these conditions are fulfilled, a recombinant clone encoding a sequence-specific DNA binding protein might be identifiable by probing protein replica filters of an expression library with radiolabeled binding site DNA. This strategy is analogous to immunological screening of an expression library and shares many of the experimental steps (Young and Davis, 1983).

Recently, gel electrophoresis DNA binding assays have been used to characterize two distinct mammalian transcription factors, H2TF1 and NF-κB, that recognize related control elements. H2TF1 binds a conserved sequence element ($^{TGGGGATTCCCCA}_{ACCCCTAAGGGGT}$) located 165 bp upstream of the initiation site of the H-2K$^b$ class I gene of the mouse major histocompatibility complex (MHC; Baldwin and Sharp, 1987; Israel et al., 1987). This sequence is present in the upstream regions of both mouse and human class I genes. The binding of H2TF1 to this site appears to stimulate MHC class I gene transcription approximately 10-fold (Kimura et al., 1986; Baldwin and Sharp, 1987). H2TF1 activity is detectable in the large variety of mammalian cell lines (mouse and human) that express MHC class I genes. Interestingly, the same control element is also recognized by NF-κB, a B–cell specific factor (Baldwin and Sharp, 1988). NF-κB was discovered by its interaction with a site in the immunoglobulin κ light chain gene enhancer (Sen and Baltimore, 1986a). NF-κB activity is restricted to B cells that transcribe the κ chain gene. Moreover, it can be induced during the differentiation of pre-B cells to B cells by treatment of the former with bacterial lipopolysaccharide (Sen and Baltimore, 1986b). This differentiation, in vitro, is accompanied by the activation of the κ gene and thus NF-κB is strongly implicated in the developmental regulation of κ gene expression (Atchison and Perry, 1987; Lenardo et al., 1987). NF-κB binding activity can also be induced in a nonlymphoid cell line (HeLa) and a T cell line (Jurkat) by treatment with a phorbol ester. In T cells, NF-κB appears to stimulate transcription of the human immunodeficiency virus (HIV) genome (Nabel and Baltimore, 1987).

Although H2TF1 and NF-κB appear to represent functionally distinct regulatory factors, their DNA recognition properties are very similar. Both factors recognize related sequences in the enhancers of the MHC class I gene, the κ chain gene, and the SV40 genome, albeit with different relative affinities (Sen and Baltimore, 1986a; Baldwin and Sharp, 1988). It is unclear whether they represent different modifications of the same protein or different proteins. Isolation of the gene or genes encoding these two activities will facilitate the analysis of their structural and functional relationships. In this study we use a DNA binding site probe, containing the conserved MHC element, to isolate a partial cDNA clone that encodes a protein whose DNA recognition properties overlap those of H2TF1 and NF-κB.

## Results

### Detection of Clones Encoding Sequence-Specific DNA Binding Proteins

The feasibility of the screening strategy was tested with a λgt11 recombinant (λEB) encoding a fusion protein that contains the DNA binding domain of the Epstein-Barr virus nuclear antigen (EBNA-1; Rawlins et al., 1985; see Experimental Procedures). Protein replica filters generated from platings of test and control phage were used to develop conditions for the specific detection of λEB plaques with a DNA probe (*oriP*) containing binding sites for EBNA-1 (data not shown; see Experimental Procedures). In reconstruction experiments, the sensitivity and fidelity of detection of λEB plaques with *oriP* DNA and α EBNA-1 antibody probes were comparable (data not shown).

A λgt11 library of cDNAs prepared with mRNA from human B cells was screened using the conditions developed with the EBNA-1 model. The DNA probe used in the screen contained a regulatory element from a mouse MHC class I gene (H-2K$^b$; Figure 1A). This sequence (MHC), along with a few flanking nucleotides, was synthesized and cloned into the pUC polylinker (see Experimen-

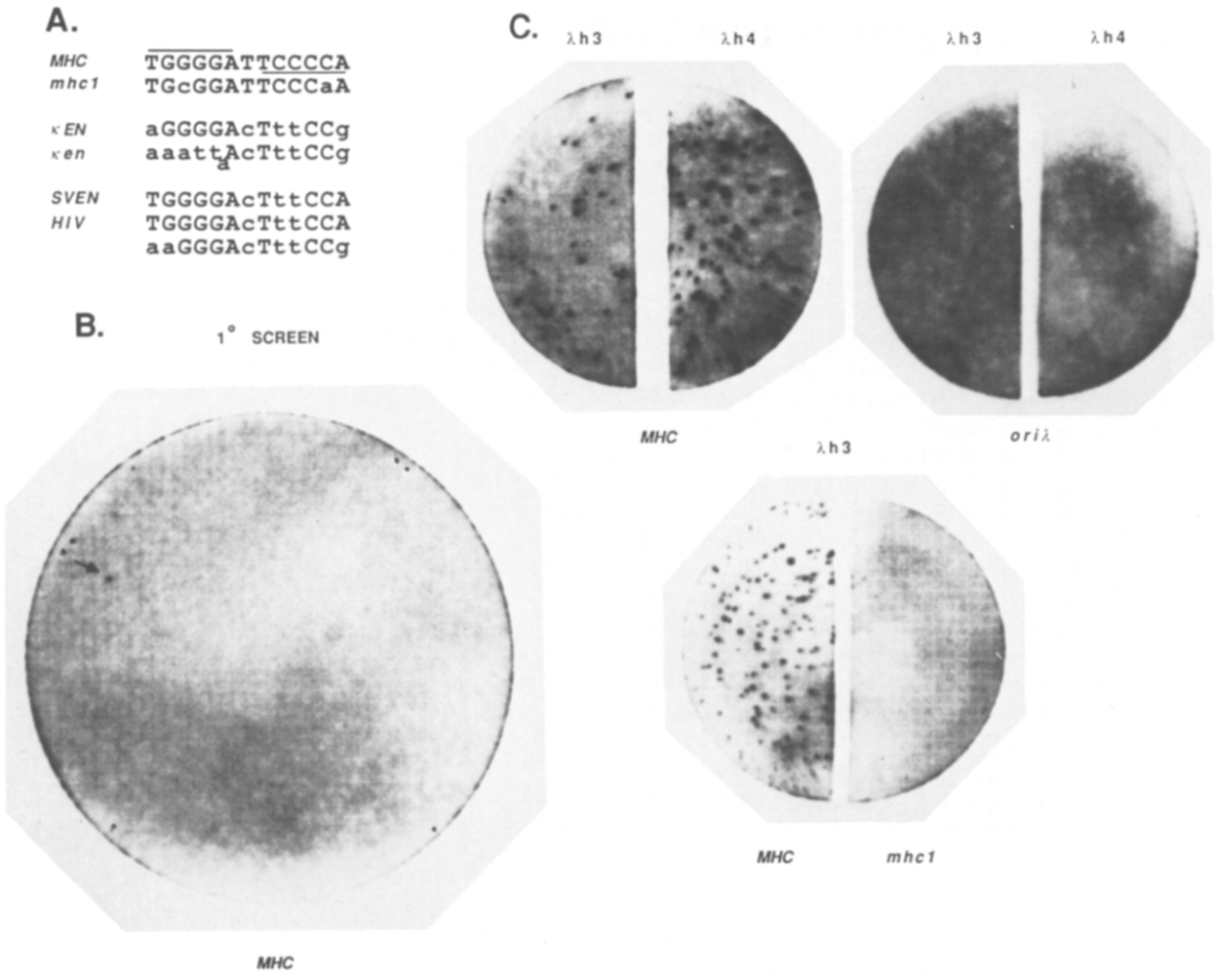


Figure 1. Isolation of λ Recombinant Clones that Specifically Bind the MHC Element

(A) H2TF1 and NF-κB binding sites. MHC is the palindromic sequence element located at position −165 in the H-2K$^b$ gene that is recognized by both H2TF1 and NF-κB. The mutant sequence mhc1 differs from the wild-type element by two transversions (G→c and C→a) and is not recognized by either factor. The κEN sequence, located in the mouse immunoglobulin κ gene enhancer, is recognized by NF-κB. The mutant sequence κen is not bound by NF-κB. The binding sites for NF-κB in the SV40 72 bp repeat and the HIV-LTR are shown for comparison.

(B) Detection of λh3 in the primary screen. Five replica filters (132 mm) from platings of 250,000 pfu of the λ cDNA library were probed with MHC DNA (2 × 10$^6$ cpm/ml). Three putative positives were detected and the one resulting in the isolation of λh3 is shown. Phage from this and another of the putative positives yielded clones that bound MHC DNA in secondary screens. The autoradiograph shown here was exposed for 24 hr at −70°C.

(C) Specificity of DNA binding by the recombinant clones. At the top, half filters generated from platings of tertiary stocks of λh3 and λh4 were probed with MHC or *oriλ* DNAs (each at 10$^6$ cpm/ml). At the bottom, half filters from a plating of plaque purified λh3 were probed with MHC and mhc1 DNAs (each at 10$^6$ cpm/ml).

tal Procedures). As discussed in the Introduction, both H2TF1 and NF-κB bind with high affinity to this MHC element. In a screen of 2.5 × 10$^5$ recombinants, two positive phage, designated λh3 and λh4, were isolated. Figure 1B shows an autoradiogram of a filter from the primary screen. The positive spot on this autoradiogram resulted in the isolation of λh3. Partially purified λh3 and λh4 phage were screened with other DNA probes to determine if their detection specifically required the MHC probe. λh3 and λh4 were not detected by the *oriλ* probe (Figure 1C). This probe contains a binding site for the bacteriophage λ O protein (Tsurimoto and Matsubara, 1981). λh3 and λh4 were also not detected by labeled pUC polylinker DNA or by a related probe (OCTA) containing a recognition site for the immunoglobulin octamer binding protein(s) (data not shown; Singh et al., 1986). A mutant MHC binding site probe (mhc1; Figure 1A) was used to test more stringently the sequence-specificity of the presumptive fusion proteins. The mutant sequence contains a transversion in each half of the symmetric MHC element and is not recognized by either H2TF1 or NF-κB (data not shown). The mhc1 probe did not detect either λh3 or λh4 plaques (Fig-

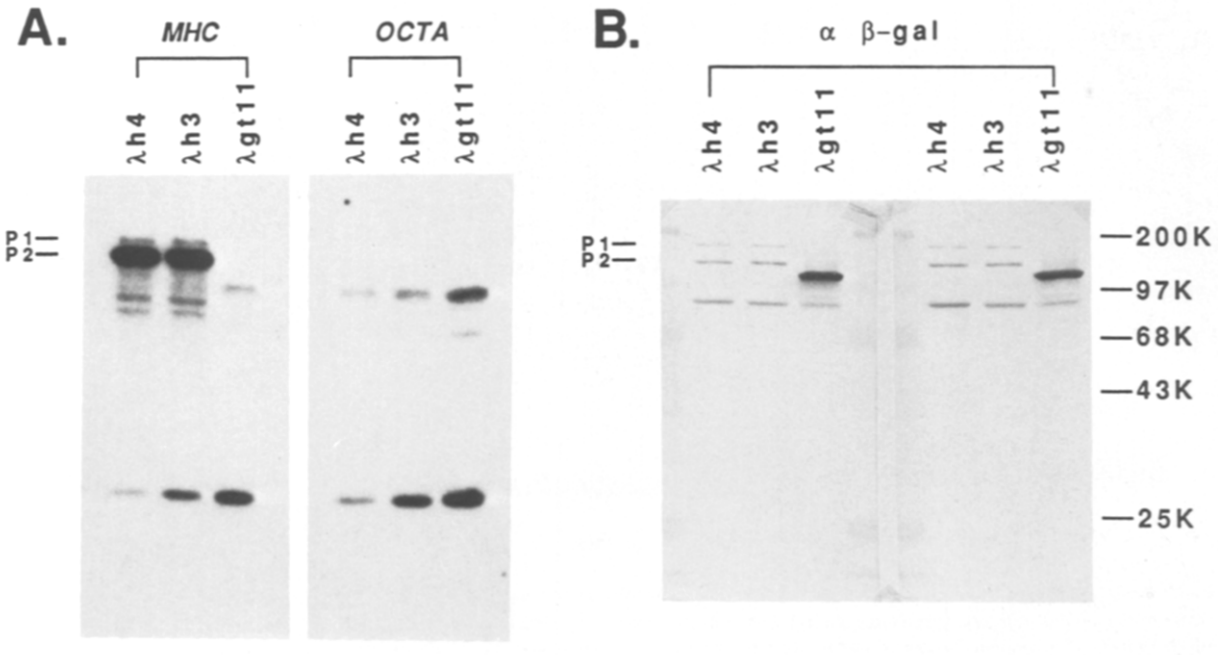


Figure 2. Western Blot Analysis of Proteins Encoded by λh3 and λh4

(A) Total proteins from induced cultures of λh4, λh3, and λgt11 lysogens were resolved by SDS-PAGE and transferred to nitrocellulose (see Experimental Procedures). Equivalent filters were probed with MHC and OCTA DNAs (each at $10^6$ cpm/ml). Bands detected with both probes represent proteins unrelated to the recombinant phage since they are present in λgt11 lanes. Furthermore, these bands do not comigrate with native β-galactosidase or any of its major cleavage products.

(B) The same filters analyzed in (A) were probed with α β-galactosidase antibodies (1:1000 dilution of rabbit serum). The two largest β-galactosidase fusion polypeptides detected in the λh3 and λh4 lanes (labeled P1 and P2) comigrated with the upper two of the four bands specifically detected with MHC probe. The positions of molecular weight standards (BRL) are indicated.

ure 1C; data for λh4 not shown). These data strongly suggested that the two λ phage express proteins that bind specifically to the MHC element.

### Characterization of the DNA Binding Proteins Encoded by λh3 and λh4

Direct evidence that the β-galactosidase fusion proteins encoded by λh3 and λh4 are responsible for the sequence-specific DNA binding activities was obtained by screening Western blots with DNA and antibody probes. Lysogens of λgt11, λh3, and λh4 were isolated and induced to generate high levels of their respective β-galactosidase proteins. Western blots of proteins from induced lysogens were prepared, and the immobilized proteins were briefly denatured with 6 M guanidine and then allowed to renature (see Experimental Procedures). This treatment increased the recovery of active molecules. Two equivalent transfers were initially probed with either the MHC element or the OCTA control DNA. A set of four bands specific to the MHC probe and the λh3, λh4 lanes was observed (Figure 2A). The two largest species of this set are labeled P1 and P2. The same transfers were then probed with antibodies to β-galactosidase. A pair of novel fusion protein bands was observed with each of the two recombinant lysogens (Figure 2B). These bands corresponded to the species P1 and P2 detected with the MHC probe. This shows that λh3 and λh4 encode β-galactosidase fusion proteins that bind specifically to the MHC element DNA. The two phage may be identical since they encode the same size fusion proteins. P1 (approximate molecular weight 160,000) probably represents the full-length fusion protein whereas P2 is a presumptive proteolytic cleavage product. Since the β-galactosidase portion of this fusion polypeptide has a molecular weight of approximately 120,000, the cDNA encoded portion must have a molecular weight of 40,000.

A gel electrophoresis DNA binding assay was used to confirm the sequence specificity of the λh3 and λh4 fusion proteins as well as to better define their recognition properties. Figure 3A shows an assay of extracts derived from the λgt11, λh3, and λh4 lysogens, with the MHC probe. A novel DNA binding activity was detected specifically in extracts of the λh3 and λh4 lysogens (complex B; lanes 4–9). This activity was IPTG-inducible, indicating that it was a product of the *lacZ* fusion gene (data not shown). A competition assay indicated that the activity represented a sequence-specific DNA binding protein (Figure 3B). Two 5′ deletion mutants of the H-2K$^b$ genomic sequence were used as competitor DNAs. The segment Δ6MHCg extends to 190 nucleotides upstream of the transcription start site and contains the MHC sequence element. The segment Δ11MHCg, on the other hand, only contains 138 nucleotides of sequences upstream of the initiation site and therefore lacks the MHC element (Baldwin and Sharp,

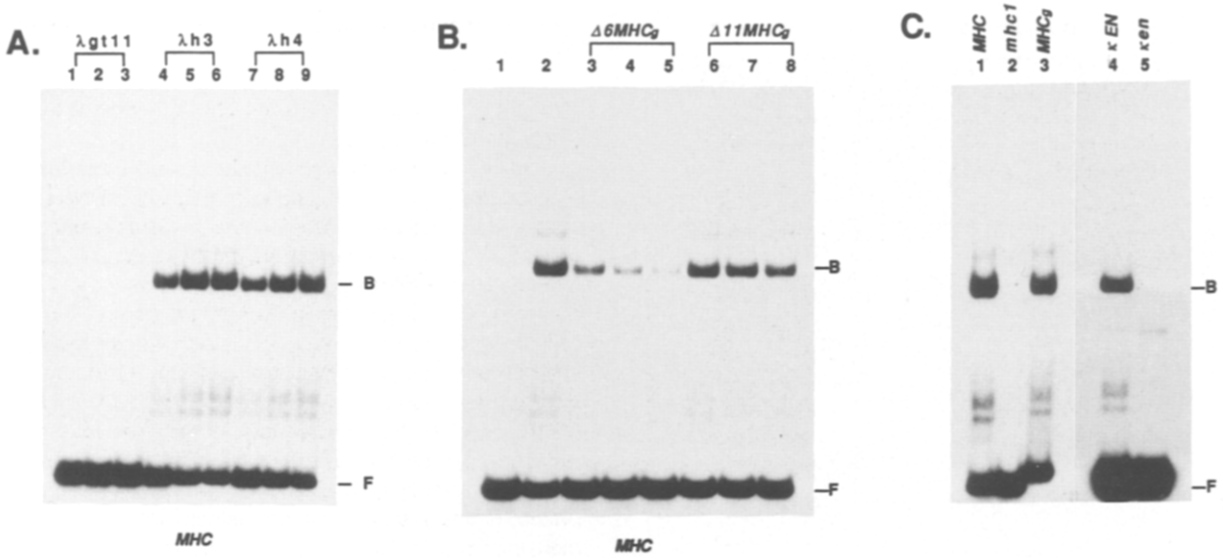


Figure 3. Gel Electrophoresis DNA Binding Assays of Proteins Encoded by λh3 and λh4

(A) Extracts prepared from induced cultures of λgt11, λh3, and λh4 lysogens were incubated with MHC DNA (25,000 cpm; 1.25 fmol) and 4 μg poly(dI-dC)·poly(dI-dC), and the reactions were resolved by electrophoresis in a nondenaturing polyacrylamide gel (see Experimental Procedures). Free and bound probes are indicated by F and B, respectively. The total protein concentrations in lanes 1, 2, and 3 were 9, 18, and 27 μg, lanes 4, 5, and 6 were 6, 12, and 18 μg, and lanes 7, 8, and 9 were 6, 12, and 18 μg.

(B) Aliquots (6 μg total protein) of an extract of the λh3 lysogen were incubated with MHC DNA as detailed in (A) in the absence or presence of varying amounts of specific-competitor DNAs. The control binding reaction (no specific-competitor DNA added) is shown in lane 2. The reactions in lanes 3, 4, and 5 contained 15, 30, and 45 fmol, respectively, of Δ6MHCg DNA. The reactions in lanes 6, 7, and 8 contained 15, 30, and 45 fmol of Δ11MHCg DNA. The free DNA probe was resolved in lane 1.

(C) Five different DNA probes were tested for complex formation with the λh3 extract. The sequences of the wild-type and mutant binding sites are shown in Figure 1A and the complete structures of the probes are described in the Experimental Procedures. In each reaction, 12 μg of total protein was added. Reactions in lanes 1 through 3 contained 25,000 cpm of each probe. Reactions in lanes 4 and 5 contained 50,000 cpm of each probe.

1987). Increasing amounts of Δ6MHCg specifically competed for the binding of the λh3 fusion protein to the MHC element oligonucleotide probe while the control Δ11MHCg did not compete (Figure 3B, lanes 2–8). It should be noted that the extended sequences flanking the MHC element in the probe used for the initial screening, the cloned oligonucleotide, are completely different from those flanking the same element in the genomic probe, Δ6MHCg. Therefore, the fusion protein appears to exclusively recognize the common MHC element. This was confirmed by a direct DNA binding assay with a genomic sequence probe (MHCg) containing the MHC element (Figure 3C, lanes 1 and 3). Both the oligonucleotide (MHC) and genomic (MHCg) probes gave rise to similarly migrating complexes. Furthermore, the double base substitution mutant (mhc1; Figure 1A) abolished recognition by the fusion protein (Figure 3C, lane 2). As previously indicated, this mutant-binding site is not recognized by either H2TF1 or NF-κB.

The immunoglobulin κ chain gene enhancer contains a binding site (κEN) for NF-κB (Sen and Baltimore, 1986a). This site is related in sequence to the MHC element, but is recognized by H2TF1 with a 10- to 20-fold lower affinity (Figure 1A; Baldwin and Sharp, 1988). A mutant κ enhancer (κEN) has been characterized both in vivo and in vitro (Lenardo et al., 1987). This mutant sequence has no B cell–specific enhancer activity and is not bound by NF-κB. The mutant contains clustered base substitutions and an insertion of a base pair in the recognition site for NF-κB (Figure 1A). The binding of the λh3 fusion protein to the wild-type κ element and the mutant version was tested (Figure 3C, lanes 4 and 5). The κEN probe generated a complex with a mobility similar to those obtained with the MHC probes. No specific complex was formed with the mutant κ enhancer DNA. Experiments in which the MHC and κ enhancer binding sites were tested for competition with binding of the MHC probe showed that the fusion protein binds with 2- to 5-fold higher affinity to the MHC site (data not shown). The κEN site differs, in part, from the MHC site by the substitution of two adenine residues for guanine residues. As discussed below, these guanine residues are probably contacted by the fusion protein.

The contacts of the fusion protein with the MHC element were probed chemically by modification of the DNA with dimethylsulfate. After partial methylation at purine residues, the modified probe was used in the gel electrophoresis DNA binding assay. Free (F) and bound (B) probe DNA was recovered, subjected to chemical cleavage at methylated sites, and the products were resolved by gel electrophoresis under denaturing conditions. Figure 4 shows the results of the methylation interference experi-

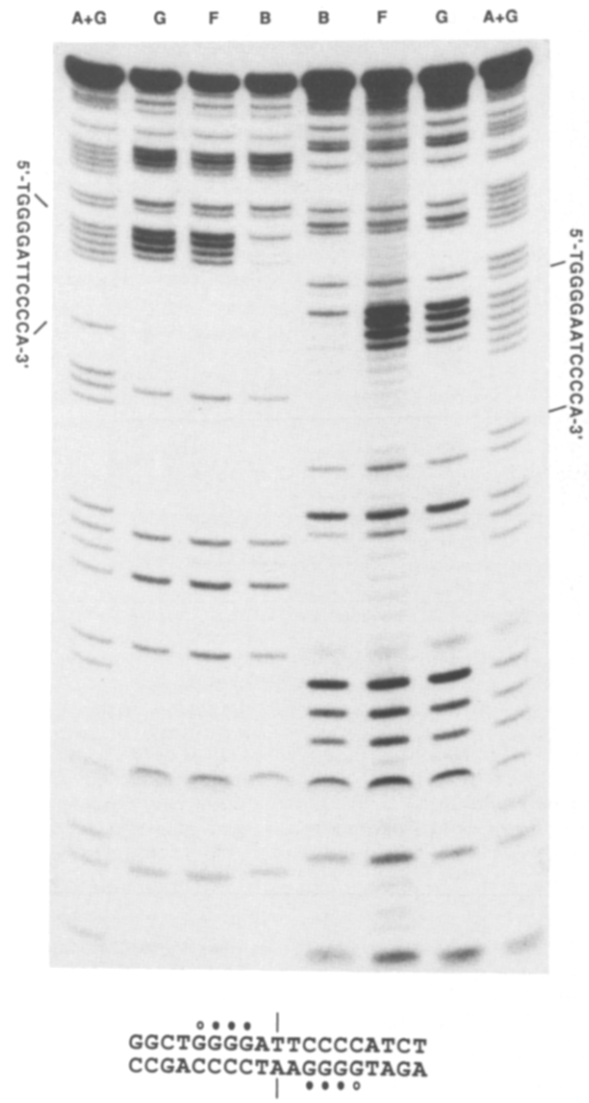


Figure 4. Methylation Interference Analysis of the Sequence Recognized by the λh3 Fusion Protein

The left four lanes show an analysis of the coding strand of the MHC probe (3′ end–labeled at the HindIII site). The right four lanes show a similar analysis of the noncoding strand (3′ end–labeled at the EcoRI site). In both cases, probe DNA was partially methylated with dimethylsulfate before addition to the binding reaction. Labeled DNA in the free (F) and bound (B) bands was recovered, cleaved at sites of methylation, and analyzed by urea-PAGE (see Experimental Procedures for details). Labeled DNAs in the reference lanes, A + G and G, were generated by cleavage at the indicated purine residues. The positions of methylated guanine residues that either strongly (●) or weakly (○) interfere with binding of the λh3 fusion protein are depicted at the bottom.

ment. On both the coding and noncoding strands, strong interference was detected when any of the internal guanine residues of each putative half site was modified at the N-7 position in the major groove. Weaker interference was observed when the external guanine residue in either putative half site was similarly modified. Thus, the fusion protein appears to symmetrically contact the MHC element in a manner similar to both H2TF1 and NF-κB (see Discussion).

### Hybridization Analysis with the cDNA Segment of the Recombinant Phage

The recombinant phage λh3 and λh4 contain cross-hybridizing and equivalent size (approximately 1 kb) cDNA segments. The inserts also have indistinguishable restriction maps and therefore appear to be identical. Southern blot hybridization confirmed that these cDNA segments are homologous to sequences in the human genome (Figure 5A). The hybridization patterns of restriction digests of various human genomic DNAs are identical. Furthermore, the fact that restriction digests with BamHI (no site in cDNA) and PstI (one site in cDNA) both generate two prominent bands suggests that the cDNAs are derived from a single copy gene. A similarly simple hybridization pattern is observed on probing restriction digests of the mouse and rat genomes (data not shown).

The expression of the human gene was analyzed by Northern blot hybridization (Figure 5B). A single, large transcript (approximately 10 kb) was observed with poly(A)$^+$ RNA from both B (X50-7) and non-B (HeLa) human cells. This transcript is moderately abundant in both cell types. The fact that the cDNA insert of the recombinant phage represents a small portion of the full-length mRNA illustrates the power of the screening strategy for the isolation of partial clones encoding sequence-specific DNA binding domains.

## Discussion

We have used a novel strategy designed to clone genes encoding sequence-specific DNA binding proteins. This strategy was undertaken with the particular aim of isolating genes specifying mammalian transcriptional regulatory proteins. The feasibility of the strategy was established by the specific detection of a phage recombinant (λEB) encoding a sequence-specific DNA binding protein (EBNA-1) with a recognition site probe. Using the conditions developed with λEB, we have screened λ phage cDNA libraries with three different DNA binding site probes. Screening with a probe containing the H2TF1 site in the MHC class I H-2K$^b$ gene led to the isolation of two identical clones that specify a putative transcriptional regulatory protein (properties discussed below). In similar screens with two other DNA probes, positive recombinant phage were also isolated at a frequency of approximately 1/100,000 (unpublished data). However, the proteins encoded by these phage appear to bind nonspecifically to either double strand or single strand DNA. Although detection of these types of clones represents an undesirable feature in screens for sequence-specific clones, their isolation suggests that recombinants encoding different types of DNA binding proteins can be detected by such functional screens of expression libraries. In future screens for recombinants encoding site-specific DNA binding proteins, the detection of these other types of clones might be selectively suppressed by inclusion of a nonspecific competitor DNA that is structurally more similar to the

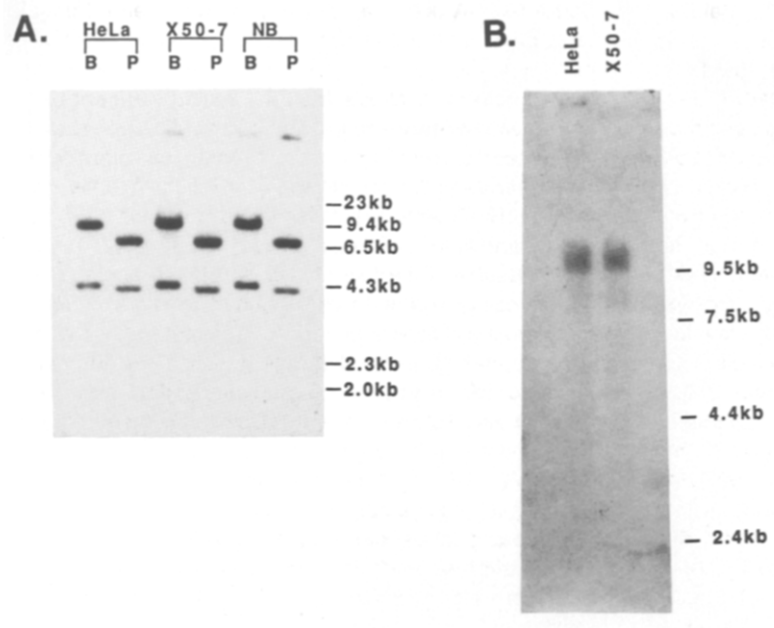


Figure 5. Southern and Northern Blot Hybridization Analysis with the λh3 cDNA

(A) Genomic DNAs (10 μg each) isolated from HeLa, X50-7 (EBV transformed human B lymphocyte), or NB (human neuroblastoma IMR32) cells were digested with either BAMHI (B) or Pst1 (P). A Southern transfer of these digests was probed with the λh3 cDNA (see Experimental Procedures). The filter was washed with 0.2× SSC at 60°C. The positions of λ HindIII standards are indicated.

(B) Cytoplasmic poly(A)$^+$ RNA (3 μg each) was prepared from HeLa and X50-7 cells and resolved in a formaldehyde-containing 1.2% agarose gel. A Northern transfer was probed and washed as above. The positions of RNA molecular weight standards (BRL) are indicated.

probe than poly(dI-dC)·poly(dI-dC). The failure to isolate sequence-specific clones with the two other DNA binding site probes may be due to the limited number of recombinants screened (~250,000 in each case), the inadequacy of the cDNA libraries, or the limitation discussed below.

Successful screening may be restricted to proteins with relatively high binding constants since only these are likely to form complexes with half-lives long enough to withstand the wash protocol. For example, if a regulatory protein has an association constant of $10^{10}$ $M^{-1}$, then under the screening conditions (the DNA probe is in excess and at a concentration of ~$10^{-10}$ M), approximately half of the active molecules on the filter will have DNA bound. Since the filters are subsequently washed for 30 min, the fraction of protein–DNA complexes that remain will be determined by their dissociation rate constant. Assuming a diffusion limited association rate constant of $10^7$ $M^{-1}$ $S^{-1}$, the dissociation rate constant in solution will be $10^{-3}$ $S^{-1}$. This rate constant translates into a half-life of approximately 10 min. Thus, one-eighth of the protein–DNA complexes should survive the 30 min wash. For a binding constant of $10^9$ $M^{-1}$, about one-tenth of the active protein molecules will have DNA bound, but virtually all of this signal should be lost since the half-life of these complexes in solution is approximately 1 min. However, it is unclear whether the equilibrium and kinetic constants of a protein–DNA interaction in solution accurately describe the binding of a DNA probe to a matrix of protein immobilized on a filter. Thus, it may be possible to isolate recombinants encoding proteins with binding constants of $10^9$ $M^{-1}$ or lower. The sensitivity of the current methodology for low affinity proteins might be significantly enhanced by using DNA probes containing multiple binding sites that are spaced so that the probe can simultaneously bind two or more immobilized protein molecules. Alternatively, it may be possible to covalently stabilize protein–DNA complexes by procedures such as UV-irradiation of preformed complexes. Since the binding constants of regulatory proteins are dependent on ionic strength, temperature, and pH, these factors might also be manipulated to enhance detection.

The λh3 recombinant expresses a β-galactosidase fusion protein that recognizes related transcription control elements in the enhancers of the MHC class I and immunoglobulin κ chain genes (see Figure 1A for sequences). This protein also binds a similar element in the SV40 enhancer 72 bp repeat (unpublished data). Furthermore, there are two putative binding sites in the long terminal repeat (LTR) of the HIV genome (Figure 1A). One of these is identical to the site in the SV40 enhancer and therefore should be recognized by the fusion protein. The existence of a clone such as λh3 was anticipated since it has previously been shown that a common factor, NF-κB, binds to the three related elements in the κ enhancer, the SV40 72 bp repeat, and the HIV-LTR (Sen and Baltimore, 1986a; Nabel and Baltimore, 1987). Interestingly, these three binding sites are more closely related to one another than they are to the MHC site (Figure 1A). The former set can be viewed as variants of the symmetric MHC site. It should be noted that the pUC polylinker contains the sequence CGGGGA, which is a variant of one of the symmetric halves (TGGGGA) of the MHC element. The fusion protein does not bind with detectable affinity to the pUC polylinker (data not shown). Thus, a high affinity interaction appears to require both symmetric halves. This conclusion is supported by the methylation interference analysis (see below).

Even though the above control elements represent quite similar sequences, they function in very different regulatory capacities. The MHC element is a component of an enhancer that functions in a variety of cell types that ex-

press MHC class I genes (Kimura et al., 1986; Baldwin and Sharp, 1987). The κ element, on the other hand, is a component of a cell-type-specific enhancer that functions only in B cells (Picard and Schaffner, 1984). The activity of this enhancer is induced in pre-B cells upon their differentiation into mature B lymphocytes (Lenardo et al., 1987). Such differentiation, in vitro, is accompanied by transcriptional activation of the κ chain gene locus (Nelson et al., 1985). The κ element appears to dictate the B-cell specificity of the κ enhancer. The different modes of functioning of the MHC and κ elements are correlated with the properties of their corresponding recognition factors, H2TF1 and NF-κB. H2TF1 activity is detected in a variety of differentiated cell types (Baldwin and Sharp, 1987; Israel et al., 1987), and this protein appears to stimulate MHC class I gene transcription approximately 10-fold (Kimura et al., 1986; Baldwin and Sharp, 1987). On the other hand, NF-κB activity is detected only in mature B cells (Sen and Baltimore, 1986a). In addition, this activity is induced during differentiation of pre-B cells to mature lymphocytes (Sen and Baltimore, 1986b). Finally, NF-κB activity is also induced by phorbol ester treatment of non-B cell lines (HeLa, Jurkat). In the case of Jurkat cells, a $T4^+$ human T cell line, NF-κB appears to stimulate the transcriptional activity of the HIV-LTR (Nabel and Baltimore, 1987). It should be noted that induction of NF-κB activity in non-B cells does not require new protein synthesis (Sen and Baltimore, 1986b). Thus the protein for NF-κB must exist in these cells before induction and be activated by a posttranslational modification.

The DNA recognition properties of the fusion protein encoded by the recombinant λh3 overlap those of H2TF1 and NF-κB. Mutants of the MHC and κ elements that are not recognized by H2TF1 or NF-κB are also not bound by the fusion protein (see Figure 1A). The recombinant protein binds the MHC element DNA with 2- to 5-fold higher affinity than the κ element. In this regard, the fusion protein has relative affinities intermediate between those of H2TF1 and NF-κB. H2TF1 binds the MHC element with 10- to 20-fold higher affinity than the κ element while NF-κB recognizes both elements with roughly equivalent affinity (Baldwin and Sharp, 1988). This intermediate relationship is also observed in the comparison of the methylation interference patterns of the three DNA-binding activities. Methylation of any of the central six guanine residues in the MHC site ($^{TGGGGATTCCCCA}_{ACCCCTAAGGGGT}$) strongly interferes with the binding of all three activities. Methylation at either of the two external guanines partially interferes with recognition by the fusion protein. In contrast, H2TF1 binding is strongly inhibited upon methylation of either of these residues while NF-κB binding shows little perturbation upon this modification (Baldwin and Sharp, 1988). This analysis of the three DNA binding activities is limited by the use of cell extracts and not purified proteins. Furthermore, the properties of a recombinant protein may be different from those of its native counterpart. Thus, it is not possible on this basis to definitively relate the protein encoded by λh3 to either H2TF1 or NF-κB.

Antibodies raised against the λh3 fusion protein will be useful in clarifying its structural relationship with H2TF1 and NF-κB. A definitive relationship will emerge from a comparison of the deduced amino acid sequence of the cDNA and the protein sequences of H2TF1 and NF-κB. It should be noted that both the H2TF1 and NF-κB proteins are present in a variety of mammalian cells and that the activity of the latter is posttranslationally regulated. Furthermore, the DNA recognition properties of these two factors are remarkably similar and overlap those of the λh3 fusion protein. These facts, as well as the observations that the cDNA in λh3 hybridizes to a single copy gene and to a single mRNA in both B and non-B cells, strengthen the possibility that all three binding activities may be products of the same gene. This hypothesis implies that H2TF1 and NF-κB represent distinct modification states of a single gene product.

## Experimental Procedures

### λgt11-EBNA-1 Recombinant

A HinfI-AhaII DNA fragment of the EBV genome (coordinates 107, 946–109, 843) that contains the EBNA-1 open reading frame was subcloned using BamHI linkers into the BamHI site of pUC13 (pUCEBNA-1; see Rawlins et al., 1985 for EBV coordinates). The λgt11-EBNA-1 recombinant was constructed by inserting the 600 bp SmaI–BamHI fragment of pUCEBNA-1 (EBV coordinates 109, 298–109, 893) into the EcoRI site of λgt11 using an EcoRI linker (GGAATTCC). A phage recombinant containing the EBNA-1 insert in the sense orientation was isolated by immunoscreening with α EBNA-1 antibodies (see below). In this recombinant, the carboxy-terminal region of EBNA-1 (191 amino acids) is fused in frame near the carboxyl terminus of β-galactosidase.

### λgt11 cDNA Expression Library

The human B cell (RPMI 4265) cDNA library constructed in the expression vector λgt11 was purchased from Clontech Laboratories, Inc. The library contains approximately $9 \times 10^5$ independent clones and has an average insert size of 1.2 kb.

### E. coli Strains

The standard pair of λgt11 host strains, Y1090 and Y1089, were employed (Young and Davis, 1983). The former was used to screen λgt11 recombinants and the latter to generate λ lysogens for the analysis of β-galactosidase fusion proteins.

### Plasmids

The plasmid pUCoriP1 was constructed by subcloning the EcoRI-NcoI fragment from the *oriP* region of the EBV genome into the SmaI site of pUC13. This fragment contains 20 high affinity binding sites for EBNA-1. pUCoriP2 was derived from pUCoriP1 by the subcloning of an *oriP* fragment (EcoRI–BstXI) of the latter into the SmaI site of pUC13. pUCoriP2 contains 11 high affinity binding sites for EBNA-1. pUCoriλ2 was made by insertion of a synthetic binding site for the bacteriophage λO protein ($^{AAATCCCCTAAAACGAGGGATAAA}_{TTTAGGGGATTTTGCTCCCTATTT}$) into the SmaI site of pUC13. The complementary oligonucleotides were a gift from R. McMacken. pUCMHCI and pUCmhcI were constructed by insertion of the following oligonucleotides: $^{GATCCGGCTGGGGATTCCCCATCT}_{\quad\quad\quad\quad GCCGACCCCTAAGGGGTAGACTAG}$ and $^{GATCCGGCTGcGGATTCCCaATCT}_{\quad\quad\quad\quad GCCGAGgCCTAAGGGtTAGACTAG}$ into the BamHI site of pUC13. The wild-type sequence is a binding site for H2TF1 and NF-κB. pUCOCTA is a similarly constructed pUC18 derivative that contains a synthetic recognition site (ATGCAAAT) for the mammalian octamer binding protein(s) (Wirth et al., 1987). The plasmids p190H2KCAT (−190 to +5) and p138H2KCAT (−138 to +5) contain 5′ deletions of the H-$2K^b$ gene upstream region fused to the coding sequence for chloramphenicol acetyl transferase (Baldwin and Sharp, 1987). All plasmid DNAs were purified by an alkaline lysis protocol followed by two bandings in CsCl ethidium bromide gradients (Maniatis et al., 1982).

### Binding Site Probes and Competitor DNAs

The MHC, mhc1, *oriλ*, and OCTA probes were generated by digesting the corresponding pUC plasmids with EcoRI and HindIII. The resulting products were end-labeled with [α-$^{32}$P]dATP using the large fragment

of E. coli DNA polymerase I (Maniatis et al., 1982). dCTP, dGTP, and dTTP were included in these reactions so as to fill in the ends of the restriction fragments. The labeled fragments were separated by native polyacrylamide gel electrophoresis. The binding site fragments (60–75 bp) were eluted from the gel and purified by ELUTIP (Schleicher and Schuell) chromatography. Using high specific activity [$\alpha$-$^{32}$P]dATP (5000 Ci/mmol), typical labelings yielded DNA probes with specific activities of 2–4 $\times$ $10^7$ cpm/pmol.

To generate the *oriP* probe, pUCoriP2 was digested with EcoRI and HindIII, and the *oriP* fragment ($\sim$400 bp) isolated by low-melt agarose gel electrophoresis. This DNA fragment was then digested with HpaII and the products labeled as detailed above. The smaller of the two HpaII fragments ($\sim$90 bp) was isolated for use as the *oriP* probe. The MHCg probe was prepared by digesting p190H2KCAT with XhoI and labeling as before. The labeled DNA was then digested with HincII and the 90 bp probe fragment purified as before. This probe contains sequence from $-190$ to $-100$ of the upstream region of the H-2K$^b$ gene. The $\kappa$EN and $\kappa$en probes were prepared according to Lenardo et al. (1987) and were gifts from P. Baeuerle.

The $\Delta$6MHCg ($-190$ to $+270$) and $\Delta$11MHCg ($-138$ to $+270$) competitor DNAs were prepared by digesting the plasmids, p190H2KCAT and p138H2KCAT, with XhoI and EcoRI. The H2KCAT fragments were isolated by low-melt agarose gel electrophoresis.

### Screening of Protein Replica Filters with DNA Binding Site Probes

Protein replica filters were prepared according to Huynh et al. (1985) with the following modifications. Plating mixtures for 100 mm and 150 mm plates contained a maximum of 2 $\times$ $10^4$ pfu and 5 $\times$ $10^4$ pfu, respectively, of $\lambda$ phage. Higher plating densities resulted in diminished signals. Infected Y1090 cells were plated in 0.7% agarose rather than agar. The IPTG saturated nitrocellulose filter overlays were incubated for 6 hr at 37°C. (Duplicate filters were not used because they appeared to be unreliable.) Plates were cooled at 4°C for 5–10 min before lifting the filters. After marking the positions of filters on plates, the filters were lifted and immediately immersed in aliquots (20 ml/82 mm filter and 50 ml/132 mm filter) of BLOTTO (5% Carnation nonfat milk powder, 50 mM Tris (pH 7.5), 50 mM NaCl, 1 mM EDTA, and 1 mM DTT; Johnson et al., 1984). In this and all subsequent steps, each filter was incubated in a separate petri plate with protein surface turned up. Filters were rapidly transferred from one solution to another to prevent drying. Filters were incubated in BLOTTO for 60 min at room temperature with gentle shaking, and washed twice (1–5 min for each wash) with aliquots (20 ml/82 mm filter and 50 ml/132 mm filter) of TNE-50 (10 mM Tris (pH 7.5), 50 mM NaCl, 1 mM EDTA, and 1 mM DTT). After these washes, the filters were either screened or stored immersed in aliquots of TNE-50 at 4°C (12–24 hr). The DNA binding activities of EBNA-1, the $\lambda$h3, and $\lambda$h4 fusion proteins were quite stable under such conditions.

For screening, the filters were incubated in aliquots (10 ml/82 mm filter and 25 ml/132 mm filter) of TNE-50 containing $^{32}$P-DNA binding site probe (1–2 $\times$ $10^6$ cpm/ml, $\sim 10^{-10}$ M) and poly(dI-dC)·poly(dI-dC) (10 $\mu$g/ml, $\sim 10^{-5}$ M bp). After 60 min at room temperature with gentle agitation, the filters were washed four times (a total of 30 min) with aliquots (20 ml/82 mm filter and 50 ml/132 mm filter) of TNE-50. They were patted dry and exposed to Kodak X-OmatAR film (12–24 hr) with an intensifying screen at $-70$°C. It should be noted that the DNA probe solution can be reused with up to five filters. Moreover, the DNA probe preparation used for screening should not contain significant levels of radiolabeled denatured strands since the latter greatly increase the background.

From a series of screens of protein replica filters generated from platings of the $\lambda$EB phage, the following conclusions were drawn: the specific detection of $\lambda$EB plaques requires a DNA probe with at least one binding site for EBNA-1 (a duplex 30-mer with a consensus binding site sequence gave a signal comparable to a probe containing two or more binding sites); DNA probes longer than 150 bp yield higher nonspecific signals; the addition of an excess of nonspecific competitor DNA [poly(dI-dC)·poly(dI-dC)] to the binding solution reduces the nonspecific signal; and both specific and nonspecific interactions of the DNA probe with proteins on the replica filter are reversible. In view of this latter point, and the fact that nonspecific interactions typically have much shorter half-lives than the specific interactions, sequence-specific binding proteins expressed to a high level can be detected after a suitable wash time.

### Screening of Protein Replica Filters with Antibody Probes

For immunological screening, protein replica filters were prepared as detailed above. They were probed with either $\alpha$ EBNA-1 antibodies (a gift from G. Milman) or $\alpha$ $\beta$-galactosidase antibodies (CAPPEL) in BLOTTO by incubating for 60 min at room temperature. After two washes with BLOTTO (5 min each), the filters were incubated (60 min at room temperature) with the above solution containing goat anti-rabbit IgG horseradish peroxidase conjugate (BIO-RAD). The immune complexes were visualized with 4-chloro-1-naphthol and $H_2O_2$ (Hawkes et al., 1982).

### Analysis of $\beta$-Galactosidase Recombinant Proteins by Western Transfers

Y1089 lysogens harboring $\lambda$gt11, $\lambda$EB, $\lambda$h3, or $\lambda$h4 phage were isolated (Huynh et al., 1985) and induced to express high levels of their respective $\beta$-galactosidase fusion proteins. Cells from 1.25 ml aliquots of the induced lysogen cultures were rapidly pelleted, and resuspended with 100 $\mu$l aliquots of SDS-PAGE loading buffer. After heating at 100°C for 5 min, proteins in each sample (25 $\mu$l) were resolved by 10% SDS-PAGE. The separation gel was then soaked in 25 mM Tris, 190 mM glycine (pH 8.3), 20% (v/v) methanol for 30–60 min. Proteins were electrophoretically transferred onto a nitrocellulose membrane (0.2 $\mu$m) using the same buffer (Towbin et al., 1979). After transfer, the filters were blocked with BLOTTO (60 min at room temperature) and then washed twice with TNE-50 (1–5 min for each wash). To assay for DNA binding activity, proteins bound to the filter were denatured and renatured according to the procedure of Celenza and Carlson (1986). The transfers were screened with DNA binding site and antibody probes as detailed for protein replica filters.

### Gel Electrophoresis DNA Binding Assays

Extracts for assaying the DNA binding activities of the $\beta$-galactosidase fusion proteins were prepared as follows: cells from 1.25 ml aliquots of the induced lysogen cultures were rapidly pelleted and resuspended in 100 $\mu$l aliquots of buffer A (50 mM Tris [pH 7.5], 1 mM EDTA, 1 mM DTT, 1 mM PMSF). Cell suspensions were subjected to a rapid freeze-thaw cycle and incubated with lysozyme (0.5 mg/ml) for 15 min on ice. NaCl was then added to 1 M and the mixtures incubated on a rotator for 15 min at 4°C. After a 30 min spin in a microfuge, the supernatants were dialyzed using Millipore filters (Type VS, 0.025 $\mu$m) against buffer A (60 min at 4°C). The dialyzed extracts were quick-frozen and stored at $-70$°C.

Binding reactions (20 $\mu$l final volume) and gel electrophoresis were performed as previously described with a modified running buffer (25 mM Tris, 190 mM glycine, 1 mM EDTA, [pH 8.3]; Singh et al., 1986). This buffer does not require recirculation during electrophoresis. The protein concentrations of the extracts were measured by the method of Bradford (1976).

### Methylation Interference Assay

The methylation interference assay was performed essentially as previously described (Staudt et al., 1986). Partially methylated MHC probes labeled uniquely at their 3′ ends (EcoRI or HindIII; 150,000 cpm) were each incubated with the $\lambda$h3 lysogen extract (45 $\mu$g protein) and poly(dI-dC)·poly(dI-dC) (10 $\mu$g) in a final volume of 50 $\mu$l. After gel electrophoresis, free DNA and protein–DNA complexes were electrophoretically transferred (in 25 mM Tris, 190 mM glycine, 1 mM EDTA, [pH 8.3]) onto NA45 DEAE membrane (Schleicher and Schuell). The wet membrane was exposed to X-ray film for 1 hr, free and bound DNA bands were cut out, and the DNA eluted according to the supplier's protocol. After cleavage with piperidine, the various DNA fractions were analyzed by urea-PAGE (10%) and autoradiography.

### Southern and Northern Blot Hybridizations

Genomic DNA was isolated from HeLa, IMR32 (human neuroblastoma), and X50-7 (EBV transformed human B lymphocyte) cells according to Maniatis et al. (1982). Restriction digests of genomic DNAs were resolved on a 1% agarose gel and transferred to Zetabind (CUNO Laboratory, Inc.).

Cytoplasmic RNA was isolated from HeLa and X50-7 cells according

to Brawerman et al. (1972). Poly(A)$^+$ RNA was purified by oligo(dT)-cellulose chromatography (Collaborative Research), separated on a formaldehyde-containing 1.2% agarose gel, and transferred to nitrocellulose.

The cDNA inserts of λh3 and λh4 were subcloned into pUC13. For hybridization analysis, the λh3 cDNA was isolated from an EcoRI digest of the pUC recombinant. A cDNA probe was generated by randomly primed synthesis with oligonucleotides (Pharmacia).

Filters were prehybridized in 55% formamide, 1× Denhardt's solution, 5× SSCPE, 0.1% SDS, 5% dextran sulfate, and 100 μg/ml denatured salmon sperm DNA for 6–8 hr at 37°C. Hybridization with $^{32}$P-cDNA (6 × $10^5$ cpm/ml) in the aforementioned solution was carried out for 14–18 hr at 37°C. Filters were washed with 0.2× SSC at 60°C. Autoradiography was performed with an intensifying screen at −70°C.


### Acknowledgments

H. S. thanks R. Hansra for encouragement throughout this work. We thank S. Speck for generously providing recombinant EBV plasmids and human cell lines. We also thank L. Chodosh for critical reading of the manuscript and M. Siafaca for patiently and carefully preparing it.

H. S. and J. H. L. acknowledge postdoctoral support from the Jane Coffin Childs Memorial Fund and the Helen Hay Whitney Foundation, respectively. A. S. B. is a Special Fellow of the Leukemia Society of America. This work was supported by grants from the National Institutes of Health (NCI PO1-CA42063), the National Science Foundation (DCB-8502718 and CDR-850003), and partially from an NCI Cancer Center Support (core) grant (P30-CA14051) to P. A. S.

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 November 4, 1987; revised December 3, 1987.


### References


Atchison, M. L., and Perry, R. P. (1987). The role of the κ enhancer and its binding factor NF-κB in the developmental regulation of κ gene transcription. Cell *48*, 121–128.

Baldwin, A. S., and Sharp, P. A. (1987). Binding of a nuclear factor to a regulatory sequence in the promoter of the mouse H-2K$^b$ class I major histocompatibility gene. Mol. Cell. Biol. *7*, 305–313.

Baldwin, A. S., and Sharp, P. A. (1988). Two factors, NF-κB and H2TF1, interact with a single regulatory sequence in the class I MHC promoter. Proc. Natl. Acad. Sci. USA *85*, pp. 723–727.

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. *72*, 248–254.

Brawerman, G. J., Mendecki, J., and Lee, S. Y. (1972). A procedure for the isolation of mammalian messenger ribonucleic acid. Biochemistry *11*, 637–641.

Celenza, J. L., and Carlson, M. (1986). A yeast gene that is essential for release from glucose repression encodes a protein kinase. Science *233*, 1175–1180.

Hawkes, R., Niday, E., and Gordon, J. (1982). A dot-immunobinding assay for monoclonal and other antibodies. Anal. Biochem. *119*, 142–147.

Huynh, T. V., Young, R. A., and Davis, R. W. (1985). Constructing and screening cDNA libraries in λgt10 and λgt11. In DNA Cloning—A Practical Approach, Volume 1, D. M. Glover, ed. (Oxford: IRL Press), pp. 49–78.

Israel, A., Kimura, A., Kieran, M., Yano, O., Kanellopoulos, J., LeBail, O., and Kourilsky, P. (1987). A common positive trans-acting factor binds to enhancer sequences in the promoters of mouse H-2 and $\beta_2$-microglobulin genes. Proc. Natl. Acad. Sci. USA *84*, 2653–2657.

Johnson, D. A., Gautsch, J. W., Sportsman, J. R., and Elder, J. H. (1984). Improved technique utilizing nonfat dry milk for analysis of proteins and nucleic acids transferred to nitrocellulose. Gene Anal. Tech. *1*, 3–8.

Kimura, A., Israël, A., Le Bail, O., and Kourilsky, P. (1986). Detailed analysis of the mouse H-2K$^b$ promoter: enhancer-like sequences and their role in the regulation of class I gene expression. Cell *44*, 261–272.

Lenardo, M., Pierce, J. W., and Baltimore, D. (1987). Protein-binding sites in Ig gene enhancers determine transcriptional activity and inducibility. Science *236*, 1573–1577.

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

Nabel, G., and Baltimore, D. (1987). An inducible transcription factor activates expression of human immunodeficiency virus in T cells. Nature *326*, 711–713.

Nelson, K. J., Kelly, D. E., and Perry, R. P. (1985). Inducible transcription of the unrearranged κ constant region locus is a common feature of pre-B cells and does not require DNA or protein synthesis. Proc. Natl. Acad. Sci. USA *82*, 5305–5309.

Picard, D., and Schaffner, W. (1984). A lymphocyte-specific enhancer in the mouse immunoglobulin κ gene. Nature *307*, 80–82.

Rawlins, D. R., Milman, G., Hayward, S. D., and Hayward, G. S. (1985). Sequence-specific DNA binding of the Epstein-Barr virus nuclear antigen (EBNA-1) to clustered sites in the plasmid maintenance region. Cell *42*, 859–868.

Sen, R., and Baltimore, D. (1986a). Multiple nuclear factors interact with the immunoglobulin enhancer sequences. Cell *46*, 705–716.

Sen, R., and Baltimore, D. (1986b). Inducibility of κ immunoglobulin enhancer-binding protein NF-κB by a posttranslational mechanism. Cell *47*, 921–928.

Singh, H., Sen, R., Baltimore, D., and Sharp, P. A. (1986). A nuclear factor that binds to a conserved sequence motif in transcriptional control elements of immunoglobulin genes. Nature *319*, 154–158.

Staudt, L. M., Singh, H., Sen, R., Wirth, T., Sharp, P. A., and Baltimore, D. (1986). A lymphoid-specific protein binding to the octamer motif of immunoglobulin genes. Nature *323*, 640–643.

Towbin, H., Staehelin, T., and Gordon, J. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc. Natl. Acad. Sci. USA *76*, 4350–4354.

Tsurimoto, T., and Matsubara, K. (1981). Purified bacteriophage λ O protein binds to four repeating sequences at the λ replication origin. Nucl. Acids Res. *9*, 1789–1799.

Wirth, T., Staudt, L., and Baltimore, D. (1987). An octamer oligonucleotide upstream of a TATA motif is sufficient for lymphoid-specific promoter activity. Nature *329*, 174–178.

Young, R. A., and Davis, R. W. (1983). Yeast RNA polymerase II genes: isolation with antibody probes. Science *222*, 778–782.