MOLECULAR AND CELLULAR BIOLOGY, June 1988, p. 2620–2627
0270-7306/88/062620-08$02.00/0

Vol. 8, No. 6

Systematic Binding Analysis of the Insulin Gene Transcription Control Region: Insulin and Immunoglobulin Enhancers Utilize Similar Transactivators

LARRY G. MOSS,* JENNIFER BARNETT MOSS, AND WILLIAM J. RUTTER

Hormone Research Institute and Department of Biochemistry and Biophysics, University of California, San Francisco, California 94143

Received 20 November 1987/Accepted 22 March 1988

The 5′ regulatory region (−345 to +1) of the rat insulin I gene (Ins-I) was examined for binding to cellular factors with short oligodeoxynucleotide probes. Over 40 binding species were detected. The binding profiles were specific for each cell type studied. We characterized the factors binding two elements crucial for enhancer activity (the Nir and Far boxes) which bear sequence similarity to the μE1, μE2, and μE3 elements of the immunoglobulin heavy-chain enhancer. The Nir box binds three cellular factors that display preferential affinities for μE1, μE2, or μE3, and the Far box binds two factors related to μE2 or μE3. The insulin gene enhancer was mutated at the Nir box element to reflect the sequences of μE1, μE2, or μE3. Ins-μE2 was fully active, Ins-μE3 was partially active, and Ins-μE1 was inactive. Thus, factors similar or identical to nuclear factor NF-μE1, NF-μE2, or NF-μE3 may play a role in the activity of the insulin gene enhancer.

In mammals, insulin gene expression is restricted to beta cells of the pancreatic islet. Transfection experiments have shown that the 5′-flanking region of the insulin gene can direct selective expression of a linked reporter function in insulin-producing cell lines (49). This region targets T-antigen expression exclusively to pancreatic beta cells in transgenic mice (20). Two cell type-specific regulatory elements appear to be operative: an enhancer (49) and a promoter (11). Cell-specific enhancers and promoters have been characterized in other genes, such as those of the immunoglobulins (2, 3, 15, 16, 19, 30, 32, 33, 37–39). In addition, sequences involved in repression of insulin gene expression in non-insulin-producing cell lines have been localized to this region (35).

Recently, the capability of the insulin 5′ regulatory elements to control expression has been analyzed by systematic replacement mutations in 10-base-pair segments (24). No single block mutation eliminates activity, but three regions independently reduce expression by 5- to 10-fold: the TATA box of the promoter and elements at −104 to −112 (Nir box) and −233 to −241 (Far box). Simultaneous mutation of these latter two elements, which have similar sequences, reduces activity to undetectable levels.

It has been presumed that *cis*-acting elements in the regulatory region exert control via interactions with *trans*-acting factors, most likely proteins (for a review, see reference 28). Binding to DNA regulatory domains has been detected by DNase I footprinting (14) or electrophoretic mobility shift assays (5, 13, 48). Both of these methods are highly sensitive to the nature of the DNA probe used, even though individual protein-binding sites may only involve a few nucleotides. Early footprinting experiments failed to reveal binding to the important Nir and Far box regulatory loci (36). In the present study, we systematically screened by shifts in electrophoretic mobility for factors that bind in a sequence-specific manner to short overlapping sequences from the insulin gene 5′ regulatory region. Significantly, most of these short domains, including the Nir and Far boxes, bind proteins in a distinctive, sequence-specific manner.

A more detailed study was carried out on the Nir and Far box regions, which bear strong sequence similarities to three of the structural motifs present in the immunoglobulin heavy-chain (IgH) enhancer (μE1, μE2, and μE3) (7, 12) (see Table 1). These E loci, along with related sequences in the kappa light-chain enhancer, bind distinct cellular factors (27, 44, 51) and act as transcriptional activation elements (27).

The three factors that bind to the Nir and Far boxes can be differentiated from one another by selectively binding different E elements as determined by competition experiments. Point mutants and IgH-insulin (InS) gene chimeras have been synthesized and tested in similar binding assays for activity in driving expression of a linked CAT (chloramphenicol acetyltransferase) reporter function (17). A positive correlation was found between the capacity of a given E motif to compete with the Nir box in a gel shift assay and its ability to stimulate transcription in transfection experiments.

MATERIALS AND METHODS

Cell culture and extracts. Hamster insulinoma (HIT-T15 M2.2.2) and fibroblast (BHK-21) cells were grown as previously described (11). All cultures contained 100 U of penicillin per ml and 100 μg of streptomycin per ml. Whole-cell extracts were prepared (29) with a final protein concentration of 5 to 7 mg/ml.

Electrophoretic mobility shift assays. The binding experiments (5, 13, 48) used double-stranded synthetic oligonucleotides as probes. Oligonucleotides were synthesized from β-cyanoethyl phosphoramidites, purified on preparative polyacrylamide gels, and eluted into water. One strand of each probe was phosphorylated with polynucleotide kinase and [γ-^{32}P]ATP (6,000 Ci/mmol) and annealed with a twofold excess of its complementary strand. This was done by heating both strands to 95°C in 10 mM Tris (pH 7.9)–2 mM $MgCl_2$–50 mM NaCl–1 mM EDTA and allowing the mixture to cool to 25°C over 2 h. Unlabeled competitor fragments

* Corresponding author.

were annealed in an identical manner. Double-stranded probes were purified by nondenaturing polyacrylamide electrophoresis, elution into 500 mM NH_4OAc–10 mM $Mg(OA_2)_2$–1 mM EDTA–0.1% sodium dodecyl sulfate at 37°C overnight, phenol-chloroform extraction, and ethanol precipitation. A ^{32}P-labeled probe (about 0.5 ng/25 fmol) was mixed with 5 μg of whole-cell extract in a total volume of 20 μl containing 10 mM *N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid (HEPES; pH 7.8), 75 mM KCl, 2.5 mM $MgCl_2$, 0.1 mM EDTA, 1 mM dithiothreitol, 10% glycerol, and 1.5 μg of poly(dI-dC):poly(dI-dC). For competition experiments, a 50-fold molar excess of competitor DNA was preincubated with the appropriate extract for 5 min at room temperature before addition of a labeled probe. After incubation at room temperature for 30 min, the binding mixtures were loaded onto 4.5% polyacrylamide gels (nondenaturing; acrylamide-bisacrylamide ratio, 30:1) after preelectrophoresis at 10 V cm^{-1} for 15 min. The gels were run at 10 V cm^{-1} in 45 mM Tris base–45 mM boric acid–1 mM EDTA. Gels were dried and autoradiographed at −70°C with an intensifier screen.

Methylation interference analysis. Essential guanine residues required for the binding of nuclear factor (NF)-InsE1 to the Nir box element were identified by the methylation interference technique (44, 45). For these experiments, a −85 to −159 rat Ins-I gene fragment, subcloned into plasmid pUC18, was asymmetrically labeled with ^{32}P-labeled nucleoside triphosphates and Klenow enzyme after digestion at either the −85 or the −159 end (*Eco*RI or *Hin*dIII) of the PUC polylinker. The labeled fragment was cleaved at the opposite end and purified on a native polyacrylamide gel as described above. The probes were methylated with dimethyl sulfate for 3 min at 18°C (31). After ethanol precipitation, 60,000 cpm of probe was mixed with 100 μg of HIT whole-cell extract and 20 μg of poly(dI-dC):poly(dI-dC) with the above-described buffer conditions in a total volume of 140 μl. To minimize the contribution of an adjacent binding site associated with probe 12.15 (see Results), 200 ng of unlabeled probe 12.15 was added to the incubation mixture as a competitor. After 30 min of incubation at room temperature, the reaction mixture was electrophoresed on a 4.5% polyacrylamide gel as described above. The bound and free fragments were identified by autoradiography for 4 h at 4°C, excised, eluted, and extracted as described above. The DNA samples were suspended in 100 μl of 1 M piperidine and cleaved at 90°C for 30 min. The products were lyophilized to dryness and twice suspended in distilled water and lyophilized. About 1,000 cpm of each sample was suspended in formamide-dye loading buffer, heated to 68°C for 5 min, and electrophoresed on a 10% polyacrylamide–8 M urea–10% formamide sequencing gel. The gels were dried and autoradiographed at −70°C with an intensifier screen.

Mutant plasmid construction. The construction of the parent plasmid pOK-1 containing −345 to +1 of the rat Ins-I gene, the entire chloramphenicol acetyltransferase (CAT)-coding sequence and the simian virus 40 (SV40) splice-polyadenylation domain are described elsewhere (24).

The technique of oligonucleotide assembly was used to construct the mutant plasmids (24). Oligonucleotides of 20 to 25 bases were prepared with 8- to 12-base-pair (bp) overlaps to reconstruct the −160 to +1 rat Ins-I sequences with *Pst*I and *Hin*dIII ends, respectively. Replacement oligonucleotides which span the region −99 to −125 were designed which contain either the stretches of IgH enhancer sequences (see Fig. 6) or point mutations (see Table 1). To facilitate cloning, the assembled construction contained a *Bam*HI site at −85 known to have no effect on activity of the 5'-flanking region (data not shown). A mutant −160 to +1 region was first assembled from component oligonucleotides that resulted in a *Pst*I end at −160 and a *Hin*dIII end at +1 as previously described (24).

The vector was prepared by excision of the native rat Ins-I −160 to +1 region with *Pst*I and *Hin*dIII (both unique), followed by dephosphorylation. The assembled, unpurified oligonucleotide construct was ligated to the vector by incubation of 100 ng of the oligonucleotide mixture with 50 ng of the vector under the above-described T4 ligase conditions. The ligation mixture was used to transfect *Escherichia coli* HB101 cells. After identification of positive colonies by restriction analysis of DNA from 2-ml overnight cultures, the DNA was further purified, and the rat Ins-I region was sequenced by a modification (24) of the double-strand method (6).

DNA transfection and expression. HIT T-15 M.2.2.2 and BHK-21 cells were cotransfected with mutant or wild-type CAT plasmids and the Rous sarcoma virus (RSV)–β-galactosidase (β-Gal) internal control plasmid by the previously described calcium phosphate technique (49).

Cell extracts prepared 48 h after transfection were analyzed for CAT and β-Gal activities (11, 49). The β-Gal values were used to normalize the individual test CAT levels (11). For RNA analysis, RSV-CAT was cotransfected with the insulin-CAT test plasmids as an internal control. RNA was isolated (11) and analyzed by reverse transcription from an oligonucleotide primer to CAT sequences (11).

RESULTS

Sequential overlapping segments of the 5'-flanking region independently bind multiple factors. The specific binding of cellular factors in hamster insulinoma whole-cell extracts to synthetic, double-stranded oligonucleotide probes was assayed by the electrophoretic mobility shift assay (5, 13, 48). However, instead of using conventional long restriction DNA fragments, we elected to use short segments whose lengths are more comparable to those of most known protein-binding regions. The 5'-flanking region of the rat Ins-I gene (−340 to −1) was subdivided into segments of 32-bp average length (except probe 1.6), with roughly 10 bp of overlap between adjacent fragments (Fig. 1). Thus, each provided about 12 bp of a unique sequence. The boundaries were in part determined by similarity with known regulatory elements, such as the TATAA box, and by the results of the previous mutational analysis (24). They are numbered by decades of the sequence, increasing in the upstream direction from the initiation of transcription. Despite the overlap between neighboring segments, each probe exhibited specific binding characteristics. In hamster insulinoma (HIT) extracts, at least 44 bands were detected among the 15 probes used to span the 340-bp 5' flank (Fig. 2A).

To probe the specificity of binding, competition experiments were conducted in the presence of a 50-fold excess of either an unlabeled probe of the test sequence (Fig. 2C) or a DNA fragment of a different sequence but of a similar length (Fig. 2B). Nearly all signals were competed for by the specific fragments. A control double-stranded oligonucleotide, which had little sequence similarity to any of the test segments, competed successfully for 17 bands. Therefore, at least 27 bands appear to result from sequence-specific binding. This nonspecific effect might be explained by partial sequence or structural similarities with the nonspecific competitor or by a relative lack of specificity of the binding reactions.

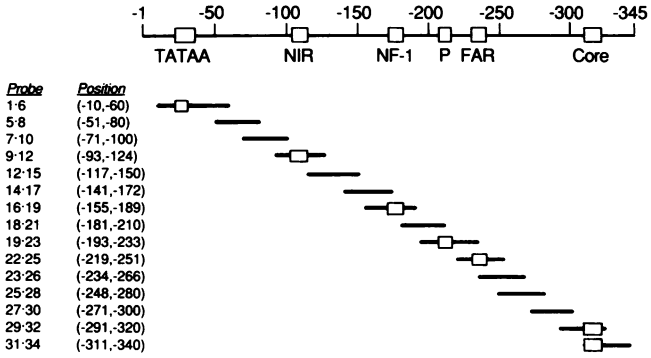


FIG. 1. Miniprobes scanning the regulatory region of the insulin gene. The probes are labeled on the basis of sequence position relative to the transcriptional start site (rounded to the nearest decade). Open rectangles indicate positions of sequences previously shown by mutagenesis analysis to be important for activity or segments related to other characterized regulatory elements (see Results). Their coordinates are as follows: TATAA, −26 to −30; Nir, −104 to −113; NF-1, −176 to −188; P (AP-1-like site), −207 to −215; Far, −230 to −239; core (SV40), −310 to −317.

It is unlikely that a significant fraction of these specific bands represents factors binding to single strands of the probe, because only the double-stranded species of free probe was observed. Also, when an excess of the unlabeled strand of the probe was added to the binding reaction, the intensities of the specific bands were unaffected and only a few faint, rapidly migrating, nonspecific bands were eliminated (data not shown). Of note is the most rapidly migrating band seen with probe 9.12 in Fig. 2B and C but not in A. It was produced as an artifact of freezing and thawing of the extract used for B and C.

The same set of probes produced a distinct binding profile with each of several cell types tested: rat insulinoma (RIN), mouse L, hamster fibroblast (BHK), COS, and HeLa (data not shown). After normalization of the migration of the bands with respect to the migration of the free probes, many bands found in the HIT cell extracts coincided with those in the other cell types, but their relative intensities varied greatly. Furthermore, several factors clearly present in insulinoma extracts were conspicuously absent from the other cell extracts.

Some of the factors that bound to the insulin DNA sequences are related to known entities (Fig. 1). Probe 1.6 (Fig. 2A) contains the TATA box and produced five intense bands; one of these may be a complex that includes transcription factor IID, a DNA-binding factor that is required for TATA box-dependent transcription (42). Probe 16.19, which has a CCAAT box sequence motif, produced the strongest single signal detected. It is likely that different factors bind to CAAT sequences (9), two of which, CCAAT box transcription factor–NF-1 (23, 41) and CAAT-binding protein C (18), have been purified. Probe 19.23 binds several specific factors in HIT cell extracts. It contains a sequence (located at −215 to −207), similar to the AP-1-binding site in the SV40 enhancer, that binds a factor in COS-7 cells which may be related to AP-1 (34). AP-1 binds not only to the SV40 enhancer but also to several eucaryotic transcriptional regulatory regions (1, 25, 26). Probe 29.32, which contains a sequence similar to the SV40 enhancer core region (TGTGGAA) (50), produced multiple bands. A core-binding factor has recently been purified from rat liver nuclei (22).

Crucial Nir and Far box loci related to immunoglobulin enhancer E motifs. The crucial related enhancer octanucleotides at −104 to −112 and −233 to −241 (the Nir box and Far box) are present in probes 9.12 and 22.25, respectively. Both probes displayed multiple binding signals, but despite their sequence similarity, the binding patterns were not identical (Fig. 2A). The two most slowly migrating bands generated by each probe appeared to be identical, but the prominent faster migrating band, evident with 9.12, was not detected with 22.25. The faint, most rapidly migrating bands seen with these and some other probes may represent nonspecific single-stranded DNA-binding factors, since they were easily competed for by various single-stranded oligonucleotides (data not shown). The three distinct bands produced by 9.12 (bands 9.12A, 9.12B, and 9.12C) were competed for by the unlabeled probe but not by the same fragment containing the inactive −104 and −112 block substitution mutation (Nir-m) (see Fig. 4A). Unlabeled probe 22.25 competed for bands 9.12A and 9.12B, but not band 9.12C, in agreement with the initial binding profiles.

Similar 32-bp oligonucleotide probes containing the mouse immunoglobulin heavy-chain enhancer sequences at the μE1 (338 to 369), μE2 (370 to 401), and μE3 (389 to 420) loci (12) were synthesized for use as competitors (Fig. 3A). Probe 9.12 was altered by point mutation at the −104 to −112 Nir box site so as to contain the cognate sequence motifs of

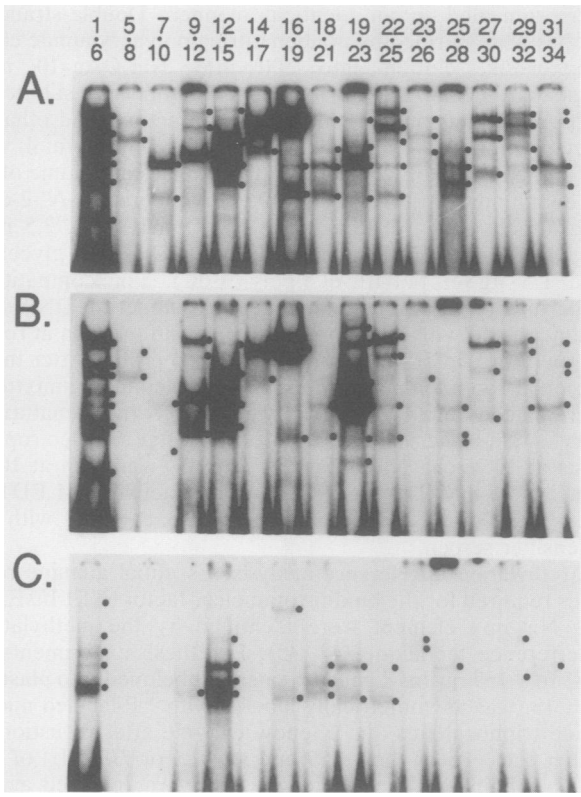


FIG. 2. Factors that bind to insulin gene sequences in HIT whole-cell extracts. (A) The probes are numbered as outlined in the legend to Fig. 1. Each probe (25 fmol) was incubated with 2 μg of poly(dI-dC):poly(dI-dC) and 5 μg of protein of the extract. The complexes were separated on a 5% nondenaturing polyacrylamide gel. The major bands are marked with dots for comparison. (B) As for A, except that 1,250 fmol of an unrelated 32-bp double-stranded oligonucleotide (5′-GATCGGACCTCGCAACGAGTATCATGGT GAGG-3′) was added to each incubation mixture as a nonspecific competitor. (C) Same conditions as for A, except that 1,250 fmol of each unlabeled probe was added to the respective incubation mixture as a specific competitor.

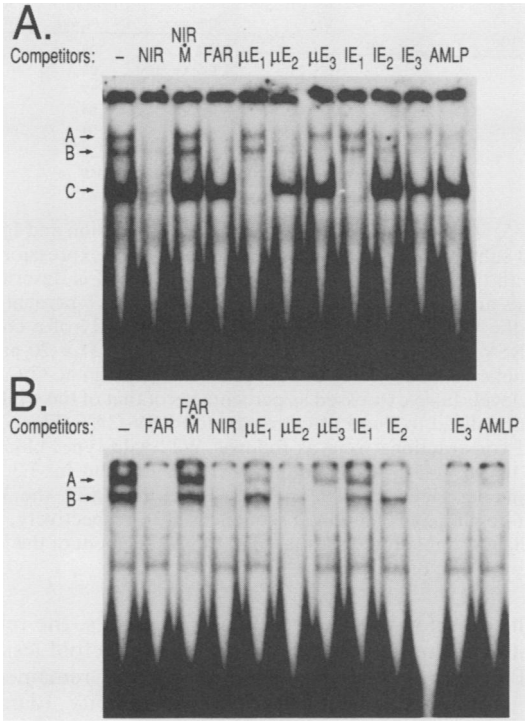


FIG. 3. Competition analysis of Nir box- and Far box-containing probes in HIT cell extracts. (A) The incubation conditions are described in the legend to Fig. 2. A 25-fmol sample of the Nir box containing ^{32}P-labeled probe 9.12 was competed for by 1,250 fmol of the following 32-bp sequences: Nir, segment 9.12; Nir-m, segment 9.12 with a −104 to −112 block transversion mutation; Far, probe 22.25; μE1, IgH enhancer (338 to 369); μE2, IgH enhancer (370 to 401); μE3, IgH enhancer (389 to 420); IgE1, IgE2, and IgE3, segment 9.12 with point mutations described in Table 1; AMLP, adenovirus major late promoter (−43 to −74). The IgH coordinates are from reference 20, and the AMLP coordinates are from reference 22. The labeled bands are 9.12A, 9.12B, and 9.12C. (B) The same conditions as for A were used, except that 22.25 was used as a probe. The competitors were as described above, except that Far-m was segment 22.25 with a −233 to −241 block transversion mutation. The labeled bands are 9.12A and 9.12B.

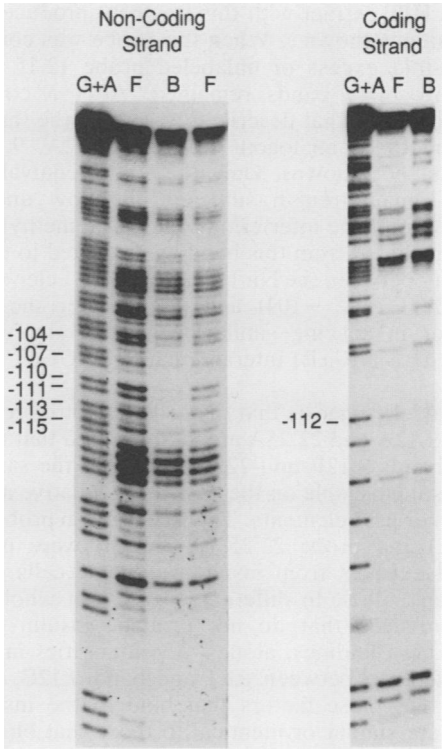


FIG. 4. Methylation interference analysis of NF-1–InsE1 binding. A partially methylated −85 to −159 insulin fragment, ^{32}P labeled on the coding or the noncoding strand (see Materials and Methods), was incubated with 200 μg of protein of HIT whole-cell extract, 20 μg of poly(dI-dC):poly(dI:dC), and 1 pmol of unlabeled probe 12.15. The major band (B) and free probe (F) were recovered, cleaved by piperidine, and resolved by electrophoresis on a 10% sequencing gel. The same unmethylated fragment was cleaved by the G+A (29) reaction for sequence position determination. Contact points are indicated at numbered sequences. The numbers to the left indicate the bases protected.

either μE1, μE2, or μE3 in the background of the native insulin sequence (abbreviated Ins-E1, Ins-E2, and Ins-E3, respectively [Table 1]). Band 9.12A was competed for by μE2 and Ins-E2; band 9.12B was competed for by μE3, Ins-E3, and to a lesser degree by μE2; and band 9.12C was competed for by μE1 and Ins-E1. Competition for band 9.12B by μE2 disappeared at lower competitor concentrations at which μE3 and Ins-E3 were still effective (data not shown).

TABLE 1. Sequence comparison of regulatory elements

Element	Sequence[a]	
Nir	−113 CGCCATCTGC	−104
Far	−239 GGCCATCTGG	−230
IgHμE1[b]	358 GGCCATCTTG	349
IgHμE2[b]	390 TGCCAGCTGC	381
IgHμE3[b]	409 TGCCACATGA	400
AMLP-USE[c]	−63 GGCCACGTGA	−54

[a] Underlined bases represent the base differences between the μE1, μE2, and μE3 sequences and the Nir box sequence that were used to generate the immunoglobulin (or Ins) E1, E2, and E3 mutants of the Nir box for competition (Fig. 3) and expression (Fig. 5) studies.
[b] Immunoglobulin heavy-chain enhancer (22).
[c] Adenovirus major late promoter upstream element (20).

A similar analysis was carried out on probe 22.25 containing the Far box (Fig. 3B). The principal bands 22.25A and 22.25B were compete for by both unlabeled probes 9.12 and 22.25 but not by the inactive block substitution mutation at −233 to −241 (Far-m). Band 22.25A was competed for by μE2 and Ins-E2, while band 22.25B was selectively competed for by μE3 and Ins-E3. μE1 and Ins-E1 failed to compete for either band. In principle, specific binding could occur at the ends or in the internal regions of the probe. It seems likely, however, that most probe interactions occur in the central region, since probes with common overlapping sequences rarely share common bands. The failure of probes containing mutated sequences in the central region of the Nir and Far boxes (probes Nir-m and Far-m, respectively) (Fig. 3) to compete for factors that bind to the wild-type probes supports this contention.

The Nir box binding site was further characterized by methylation interference studies (Fig. 4) which delineated guanine residues whose methylation precludes factor binding, presumably at the point of contact between protein and DNA (45). A somewhat longer fragment was needed to delineate essential residues clearly on a sequencing gel. For this purpose, we chose a 74-bp (−85 to −159) *Bam*HI-*Pst*I fragment containing the Nir box element. A mobility shift

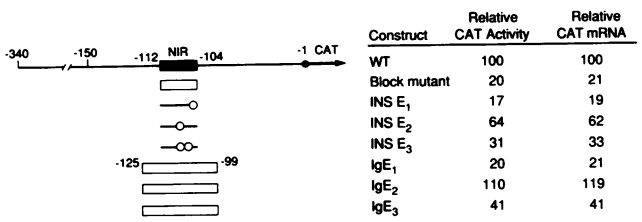
assay of a HIT extract with this fragment produced several bands (data not shown). When this probe was coincubated with a 50-fold excess of unlabeled probe 12.15 (−117 to −150), only three bands remained. After a competition analysis similar to that described by Fig. 3, the three bands were found to be analogous to bands 9.12A, 9.12B, and 9.12C (data not shown). Only the 9.12C equivalent band provided signal strength sufficient to allow unequivocal determination of the interfering bases. The methylated guanines (*) excluded from this band were limited to the region of similarity between the Nir box and the μE1 element (−115 CTCGCCATCTGC −104) and were either the identical residues or in spacing similar to that of the previously described μE1–NF-μE1 interaction (360 TCGGCCATCTTG 349) (51).

This analysis suggests that the cellular factors responsible for bands 9.12A and 22.25A are identical and that those that make up bands 9.12B and 22.25B are also the same. Each band is distinguishable on the basis of its relative affinity for μE1, μE2, or μE3 elements. The signals from probe 9.12 (A, B, and C) and probe 22.25 (A and B) were present in whole-cell extracts from insulin-producing cells and were also present, albeit to different degrees, in whole-cell extracts from cells that do not produce insulin (data not shown). These findings, along with similarities in methylation interference between μE1 and band 9.12C, led us to conclude that these factors that bind to the insulin gene enhancer are similar or identical to those that bind to their related sequences in the IgH enhancer. We therefore refer to band 9.12C as NF-InsE1, bands 9.12A and 22.25A as NF-InsE2, and bands 9.12B and 22.25B as NF-InsE3.

The sequences of μE3 are related to the upstream activating element of the adenovirus late promoter (−54 to −63) (21) (Table 1). This sequence also competes for bands 9.12B and 22.25B just as μE3 and Ins-E3 do (Fig. 3). It has recently been shown that this adenovirus upstream element is the binding site of a purified cellular protein (major late transcription factor) (8).

IgH μE2 element substituted for the Nir box is fully active. The biological activity of various mutations of the Nir box regions was tested by measuring the effects of substitutions of appropriate segments in the insulin 5′-flanking region on the transient expression of linked CAT sequences in HIT cells (Fig. 5). The Nir box was selected as the site for mutation, since the neighboring sequences are relatively insensitive to mutation (24). We used the overlapping oligonucleotide ligation technique used in our systematic block mutagenesis studies (24). In parallel with our binding studies, 27 bp of an IgH gene enhancer sequence containing either the native μE1, μE2, or μE3 element was substituted for the −99 to −125 segment of the insulin gene (Fig. 5). Alternatively, substitutions were made in the −104 to −112 region (Table 1) to produce a μE1, μE2, or μE3 analog (Ins-E1, Ins-E2, or Ins-E3 [Fig. 5]). In each instance, the distance relationships between the cognate E motifs and remaining native insulin gene sequences were maintained in the wild-type configuration.

The activity of these constructs in transfection experiments with HIT cells was measured by CAT activity and by primer extension (Fig. 5 and 6). μE1 and Ins-E1 exhibited activity similar to that of the block transversion mutation of the Nir box and are therefore considered inactive. However, μE3 and Ins-E3 showed significant activity, about twice the activity of the block mutant. In contrast, the Ins-E2 construct approached wild-type activity, and the μE2 plasmid displayed full wild-type activity. When this analysis was performed in non-insulin-producing BHK cells, the relative CAT activity for all constructs was about 200-fold less than in HIT cells, but the rank order of activity remained the same, with μE2 exhibiting wild-type activity (data not shown). The simplest interpretation of these experiments is that binding of the NF-InsE2 factor is associated with the positive transcriptional effect of the Nir box element, whereas binding with NF-InsE1 yields no activation.



Construct	Relative CAT Activity	Relative CAT mRNA
WT	100	100
Block mutant	20	21
INS E_1	17	19
INS E_2	64	62
INS E_3	31	33
IgE_1	20	21
IgE_2	110	119
IgE_3	41	41

FIG. 5. Transcriptional activity of Nir box mutation and IgH μE element substitutions. Construction of insulin-CAT expression plasmids with the mutations and substitutions depicted is described in Materials and Methods. CAT activity (expressed as a percentage of that of the wild type) was normalized to the activity of a cotransfected RSV-β-Gal plasmid. A relative activity of 100 = 20 pmol of chloramphenicol acetylated per μg of extract per min at 37°C. CAT mRNA levels (also expressed as percentages of that of the wild type) were normalized to those of cotransfected RSV-CAT (Fig. 6). The plasmid constructions were as follows: WT, wild type; block mutant, −104 to −112 transversion; Ins-E1, Ins-E2, and Ins-E3, point mutations described in Table 1; IgE1, IgE2, and IgE3, the 344-to-369, 376-to-401, and 395-to-420 sequences (12), respectively, of the IgH enhancer substituted for the −125 to −99 segment of the insulin gene control region.

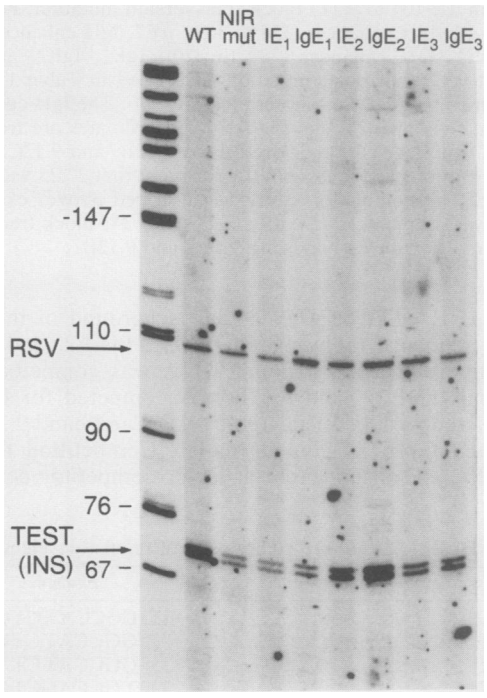


FIG. 6. Primer extension analysis of the Nir box mutations and IgH μE element substitutions depicted in Fig. 5. The plasmids are those described in Fig. 5 (IE1, IE2, and IE3 represent Ins-E1, Ins-E2, and Ins-E3, respectively). The extension products from a CAT primer (11) are 107 nucleotides for the RSV internal control and 71 nucleotides for the test plasmids. The left lane contained radiolabeled size markers. Each reaction represents 5 μg of total RNA. The numbers on the left indicate the bases protected.

DISCUSSION

Several methods have been used in attempts to detect factors that interact with the *cis*-acting regions that control insulin gene transcription. Footprinting studies in cellular extracts show multiple, broad, diffuse binding domains that cover most of the 5'-flanking region (E. Fodor and W. J. Rutter, unpublished data). Ohlsson and Edlund (36) were able to detect three binding domains in the region upstream from the promoter in nuclear extracts of insulinoma cells. However, none of these domains coincided with the Nir and Far boxes.

The electrophoretic gel shift assay provides a sensitive means for detecting minute quantities of protein-DNA complexes. We and others have found that such assays are highly dependent on the nature of the probe used and the conditions of the experiment. Since binding domains of most DNA-binding proteins are small (often less than 10 nucleotides), a probe containing extended DNA sequences can include multiple binding sites and, therefore, generate complex, difficult-to-interpret DNA-protein reactions. We thus elected to use short oligonucleotides (miniprobes) that are long enough to contain most binding sites and yet still preserve their double-stranded character. By using probes that span the entire insulin gene transcription control region, an unexpectedly large number of bands was resolved from insulinoma cell extracts (Fig. 2). The sequence specificity of most interactions was apparent from both the distinctive banding patterns of each probe and the results of differential competition experiments with specific and nonspecific DNA segments (Fig. 2). The multiple bands detected with many probes need not represent multiple, discrete DNA-binding factors, since protein-protein interactions with a single DNA-binding species may yield additional, more slowly migrating bands.

The differences in the binding profiles obtained with the miniprobes as compared with larger restriction fragments may reflect differences in concentrations and affinities of various factors or perhaps significant interactions between neighboring binding events. This is illustrated by mobility shift experiments with a long restriction fragment that comprises the core insulin enhancer (−104 to −250). Protein binding to this fragment was dominated by an interaction with the NF-1–CCAAT boxlike site centered at −180 (Fodor and Rutter, unpublished), which we detected with miniprobe 16.19 (Fig. 2). When this site was deleted from the probe (Δ118 to 191), the relatively weaker signals due to the Nir and Far boxes, as well as the −200 to −230 region, became apparent (data not shown).

As a screening test for factor binding in crude extracts, miniprobe assays are preferable to footprinting experiments because of higher sensitivity for detection of low-abundance factors, more effective resolution of binding sites, and the ability to discriminate multiple binding factors that interact at single or overlapping loci. For example, because the protection of a DNA segment requires a molar excess of protein over the DNA-binding site, the binding due to many low-abundance species seen in the miniprobe analysis would not be recognized as a footprint without significant chromatographic fractionation of extracts. Whether all of the greater than 40 components visualized in this study are operationally significant is an unresolved question. However, since a sensitive screening test seeks initially to minimize false-negative results, a small fraction of false-positive results may be expected. In vitro and in vivo reconstruction experiments will likely be necessary to define the functional relationships of these entities.

All probes that contain sequences shown by in vivo mutational analysis (24) to have regulatory significance exhibited binding to components in HIT cell extracts. The functionally important, structurally similar Nir and Far box elements bound multiple factors. The two factors that bound to the Far box appeared to be identical to two of the three factors that bound to the Nir box. The loss of activity due to the simultaneous mutation of the Nir and Far boxes was much greater than the sum of the effects seen as a result of their individual mutations (24). These elements may therefore perform symmetrical functions, with each to some extent able to compensate for loss of the other. This may be explained by our finding that these elements bound two common factors, at least one of which appeared to be essential for expression of the insulin gene.

The Nir box and Far box elements have a high degree of sequence similarity with the μE1, μE2, and μE3 regulatory elements of the immunoglobulin heavy-chain enhancer (Table 1). Specific factors which bind to the μE elements have been detected in both lymphoid and nonlymphoid cells (NF-μE1, NF-μE2, and NF-μE3) (27, 44, 51). We showed that each Nir box band can be identified by specific competition with μE1, μE2, or μE3. The two Far box factors are competed for by μE2 and μE3, respectively (Fig. 3). Since these factors have similar specificity but are not necessarily identical molecules, we term them NF-InsE1, NF-InsE2, and NF-InsE3. By virtue of their apparent presence in whole-cell extracts from multiple cell types and their participation in the regulation of different, often unrelated genes, the NF-1–Ins(μ)E factors may be analogous to the recently purified transcription factors Sp-1 and AP-1 (1, 4, 10, 25, 26).

In lymphoid cells, μE1, μE2, and μE3 serve as positive activation elements for the IgH enhancer (27). The modest activity of this enhancer in nonlymphoid cells has been attributed in part to the μE3 element (27). To test the role of these elements in the insulin gene control region, we substituted μE1, μE2, and μE3 segments for the Nir box element in the insulin gene enhancer. The μE2 construct produced full wild-type activity, whereas μE3 was slightly active and μE1 was not active (Fig. 5). Thus, it appears that NF-insE2 is primarily responsible for positive activation of the Nir box. NF-InsE1 and NF-InsE3 may, as competitors, attenuate that activity. These factors which bind to individual sites in the IgH enhancer may, in the insulin gene, provide the basis for a complex differential control mechanism that acts on a single regulatory element. The octamer sequence of the immunoglobulin kappa chain promoter is also a site for multiple-factor binding (47). Lymphoid cells, as well as many other cell types, produce a factor (nuclear factor A-1) which binds to this octamer locus and apparently activates a similar octamer element in the histone H2B promoter (46). A second factor (nuclear factor A-2), found only in lymphoid cells, binds to the same kappa octamer sequence and is a major determinant of cell-specific promoter activity (27, 47).

The evidence for multiple tissue-specific binding profiles suggests that transcriptional specificity in differentiated cells emanates from a unique combinatorial array of regulatory elements in the DNA which interact with the particular ensemble of binding factors, many of which are not necessarily cell specific. Hints at this possibility have arisen from work showing that multiple copies of individual elements of the SV40 enhancer create new enhancers that exhibit different cell specificities (43). More provocative is the de novo creation of pancreatic tissue preference in an infectious

polyomavirus by combination of its enhancer with the 72-base repeat of the Moloney murine leukemia virus enhancer (40). It is possible that the molecular context influences specificity in insulin gene expression. For example, why is NF-μ(Ins)E1 active when bound to the IgH enhancer in lymphoid cells and inactive when bound to the insulin enhancer in insulinoma cells? Although the factors may be different (by sequence or modification), the unique molecular arrangement of each gene in different cell types may be a critical determinant of NF-μE1 activity. Furthermore, despite the ubiquitous presence of NF-μE1, NF-μE2, and NF-μE3 as determined by in vitro binding, in vivo footprinting experiments suggest that binding to the IgH enhancer occurs specifically in lymphoid cells (7, 12). Therefore, other higher-order molecular interactions may also play important roles in dictating cell specificity. Whether regulation of insulin gene expression involves cell-specific factors, perhaps acting as molecular foci for a unique regulatory array, is unknown. One such candidate appears in insulin-producing cells (HIT and RIN) as a prominent band produced by probe 19.23; this signal has not been detected in extracts from other cell types tested (Fig. 2; data not shown). In addition, while NF-InsE2 has been detected in whole-cell extracts from non-insulin-producing cell types, it appears to be selectively enriched in extracts from HIT cell nucleic and absent from the nuclei of non-insulin-producing cells (data not shown). This factor may be structurally unique in HIT cells or targeted to nuclei in a cell-specific manner. More comprehensive structural analysis and, eventually, functional reconstitution with purified factors may provide the answer.

ACKNOWLEDGMENTS

We thank Michael Walker, Uri Nir, Danielle Melloul, Olof Karlsson, and the other members of the Hormone Research Institute for helpful discussions, Mickey Urdea for oligonucleotides, Jennifer Turner for plasmid production, and Leslie Spector for efficient manuscript preparation.

This work was supported by Public Health Service grant AM21344 from the National Institutes of Health and a grant from the March of Dimes, both to W.J.R.

LITERATURE CITED

1. **Angel, P., M. Imagawa, R. Chiu, B. Stein, R. J. Imbra, H. J. Rahmsdorf, C. Jonat, P. Herrlich, and M. Karin.** 1987. Phorbol ester-inducible genes contain a common *cis* element recognized by a TPA-modulated *trans*-acting factor. Cell **49**:729–739.
2. **Banerji, J., L. Olson, and W. Schaffner.** 1983. A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. Cell **33**:729–740.
3. **Bergman, Y., D. Rice, R. Grosschedl, and D. Baltimore.** 1984. Two regulatory elements for kappa immunoglobulin gene expression. Proc. Natl. Acad. Sci. USA **81**:7041–7045.
4. **Briggs, M. R., J. T. Kadonaga, S. P. Bell, and R. Tjian.** 1986. Purification and biochemical characterization of the promoter-specific transcription factor, Sp-1. Science **234**:47–52.
5. **Carthew, R. W., L. A. Chodosh, and P. A. Sharp.** 1985. An RNA polymerase II transcription factor binds to an upstream element in the adenovirus major late promoter. Cell **43**:439–448.
6. **Chen, E. Y., and P. Seeburg.** 1985. Supercoil sequencing: a fast and simple method for sequencing plasmid DNA. DNA **4**:165–170.
7. **Church, G. M., A. Ephrussi, W. Gilbert, and S. Tonegawa.** 1985. Cell type specific contacts to immunoglobulin enhancers in nuclei. Nature (London) **313**:798–801.
8. **Chodosh, L. A., R. W. Carthew, and P. A. Sharp.** 1986. A single polypeptide possesses the binding and transcription activities of the adenovirus major late transcription factor. Mol. Cell. Biol. **6**:4723–4733.
9. **Dorn, A., J. Bollekens, A. Staub, C. Benoist, and D. Mathis.** 1987. A multiplicity of CCAAT box-binding proteins. Cell **50**: 863–872.
10. **Dynan, W. S., S. Sazer, R. Tjian, and R. T. Schimke.** 1986. Transcription factor Sp-1 recognizes a DNA sequence in the mouse dihydrofolate reductase promoter. Nature (London) **319**: 246–248.
11. **Edlund, T., M. D. Walker, P. J. Barr, and W. J. Rutter,** 1985. Cell specific expression of the rat insulin gene: evidence for role of two distinct 5′ flanking elements. Science **230**:912–916.
12. **Ephrussi, A., G. M. Church, S. Tonegawa, and W. Gilbert.** 1985. B lineage-specific interactions of an immunoglobulin enhancer with cellular factors in vivo. Science **227**:134–140.
13. **Fried, M., and D. M. Crothers.** 1981. Equilibria and kinetics of *lac* repressor operator interactions by polyacrylamide gel electrophoresis. Nucleic Acids Res. **9**:6505–6525.
14. **Galas, D., and A. Schmitz.** 1978. DNase footprinting: a simple method for detection of protein-DNA binding specificity. Nucleic Acids Res. **5**:3157–3170.
15. **Gillies, S. D., S. L. Morrison, V. T. Oi, and S. Tonegawa.** 1983. A tissue specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. Cell **33**:717–728.
16. **Gopal, V. T., T. Shimada, A. W. Baur, and A. W. Nienhuis.** 1985. Contribution of promoter to tissue-specific expression of the mouse immunoglobulin kappa gene. Science **229**:1102–1104.
17. **Gorman, C. M., L. F. Moffat, and B. H. Howard.** 1982. Recombinant genomes which express choramphenicol acetyltransferase in mammalian cells. Mol. Cell. Biol. **2**:1044–1051.
18. **Graves, B. J., P. F. Johnson, and S. L. McKnight.** 1986. Homologous recognition of a promoter domain common to the murine sarcoma virus long terminal repeat and the herpes simplex virus *tk* gene. Cell **44**:565–576.
19. **Grosschedl, R., and D. Baltimore.** 1985. Cell-type specificity of immunoglobulin gene expression is regulated by at least three DNA sequence elements. Cell **41**:885–897.
20. **Hanahan, D.** 1985. Heritable formation of pancreatic B cell tumors in transgenic mice expressing recombinant insulin/SV40 oncogenes. Nature (London) **315**:115–122.
21. **Hen, R., P. Sassone-Corsi, J. Corden, M. P. Gaub, and P. Chambon.** 1982. Sequences upstream from the TATA box are required *in vivo* and *in vitro* for efficient transcription from the adenovirus serotype 2 major late promoter. Proc. Natl. Acad. Sci. USA **79**:7132–7136.
22. **Johnson, P. F., W. H. Landschulz, B. J. Graves, and S. L. McKnight.** 1987. Identification of a rat liver nuclear protein that binds to the enhancer core element of three animal viruses. Genes Dev. **1**:133–146.
23. **Jones, K. A., J. T. Kadonaga, P. J. Rosenfeld, T. J. Kelly, and R. Tjian.** 1987. A cellular DNA-binding protein that activates eukaryotic transcription and DNA replication. Cell **48**:79–89.
24. **Karlsson, O., T. Edlund, J. B. Moss, W. J. Rutter, and M. D. Walker.** 1987. A mutational analysis of the insulin gene transcription control region: expression in beta cells is dependent on two related sequences within the enhancer. Proc. Natl. Acad. Sci. USA **84**:8819–8823.
25. **Lee, W., A. Haslinger, M. Karin, and R. Tjian.** 1987. Activation of transcription by two factors that bind promoter and enhancer sequences of the human metallothionein gene and SV40. Nature (London) **325**:368–372.
26. **Lee, W., P. Mitchell, and R. Tjian.** 1987. Purified transcription factor AP-1 interacts with TPA-inducible enhancer elements. Cell **49**:741–752.
27. **Lenardo, M., J. W. Pierce, and D. Baltimore.** 1987. Protein-binding sites in immunoglobulin gene enhancers determine transcriptional activity and inducibility. Science **236**:1573–1577.
28. **Maniatis, T., S. Goodbourn, and J. A. Fischer.** 1987. Regulation of inducible and tissue-specific gene expression. Science **236**: 1237–1244.
29. **Manley, J. L., A. Fire, A. Cano, P. A. Sharp, and M. L. Gefter.**

1980. DNA dependent transcription of adenovirus genes in a soluble whole cell extract. Proc. Natl. Acad. Sci. USA **77**:3855–3859.

30. **Mason, J. O., G. T. Williams, and M. S. Neuberger.** 1985. Transcription cell type specificity is conferred by an immunoglobulin VH gene promoter that includes a functional consensus sequence. Cell **41**:479–487.
31. **Maxam, A., and W. Gilbert.** 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. USA **74**:560–564.
32. **Mercola, M., X.-F. Wang, J. Olsen, and K. Calame.** 1983. Transcriptional elements in the mouse immunoglobulin heavy chain locus. Science **221**:663–665.
33. **Neuberger, M. S.** 1983. Expression and regulation of immunoglobulin heavy chain gene transfected into lymphoid cells. EMBO J. **2**:1373–1378.
34. **Nir, U., E. Fodor, and W. J. Rutter.** 1988. Capturing nuclear sequence-specific DNA-binding proteins by using simian virus-derived minichromosomes. Mol. Cell. Biol. **8**:982–987.
35. **Nir, U., M. D. Walker, and W. J. Rutter.** 1986. Regulation of rat insulin 1 gene expression: evidence for negative regulation in nonpancreatic cells. Proc. Natl. Acad. Sci. USA **83**:3180–3184.
36. **Ohlsson, H., and T. Edlund.** 1986. Sequence-specific interactions of nuclear factors with the insulin gene enhancer. Cell **45**: 35–44.
37. **Picard, D., and W. Schaffner.** 1984. A lymphocyte specific enhancer in the mouse immunoglobulin kappa gene. Nature (London) **307**:80–82.
38. **Picard, D., and W. Schaffner.** 1985. Cell type preference of immunoglobulin kappa and lambda gene promoters. EMBO J. **4**:2831–2838.
39. **Queen, C., and D. Baltimore.** 1983. Immunoglobulin gene transcription is activated by downstream sequence elements. Cell **33**:741–748.
40. **Rochford, R., B. A. Campbell, and L. P. Villarreal.** 1987. A pancreas specificity results from the combination of polyomavirus and Moloney murine leukemia virus enhancer. Proc. Natl. Acad. Sci. USA **84**:449–453.
41. **Rosenfeld, P. J., and T. J. Kelly.** 1986. Purification of NF-I by DNA recognition site affinity chromatography. J. Biol. Chem. **261**:1398–1408.
42. **Sawadogo, M., and R. G. Roeder.** 1985. Interaction of a gene specific transcription factor with the adenovirus major late promoter upstream of the TATA box region. Cell **43**:165–175.
43. **Schirm, S., J. Jiricny, and W. Schaffner.** 1987. The SV40 enhancer can be dissected into multiple segments each with a different cell type specificity. Genes Dev. **1**:65–74.
44. **Sen, R., and D. Baltimore.** 1986. Multiple nuclear factors interact with the immunoglobulin enhancer sequences. Cell **46**:705–716.
45. **Siebenlist, U., and W. Gilbert.** 1980. Contacts between E. coli RNA polymerase and an early promoter of phage T7. Proc. Natl. Acad. Sci. USA **77**:122–126.
46. **Sive, H. L., N. Heintz, and R. G. Roeder.** 1986. Multiple sequence elements are required for maximal in vitro transcription of a human histone H2B gene. Mol. Cell. Biol. **6**:3329–3340.
47. **Staudt, L. M., H. Singh, R. Sen, T. Wirth, P. A. Sharp, and D. Baltimore.** 1986. A lymphoid-specific protein binding to the octamer motif of immunoglobulin genes. Nature (London) **323**: 640–643.
48. **Strauss, F., and A. Varshavsky.** 1984. A protein binds to a satellite DNA repeat at three specific sites that would be brought into mutual proximity by DNA folding in the nucleosome. Cell **37**:889–901.
49. **Walker, M. D., T. Edlund, A. M. Boulet, and W. J. Rutter.** 1983. Cell specific expression controlled by the 5′-flanking region of insulin and chymotrypsin genes. Nature (London) **306**:557–561.
50. **Weiher, H., M. Konig, and P. Gruss.** 1983. Multiple point mutations affecting the simian virus 40 enhancer. Science **219**: 626–631.
51. **Weinberger, J., D. Baltimore, and P. A. Sharp.** 1986. Distinct factors bind to apparently homologous sequences in the immunoglobulin heavy chain enhancer. Nature (London) **322**:846–848.