Proc. Natl. Acad. Sci. USA
Vol. 85, pp. 4228–4231, June 1988
Biochemistry

A beta-cell-specific protein binds to the two major regulatory sequences of the insulin gene enhancer

(tissue specificity/trans-acting factor/insulin enhancer binding factor 1)

HELENA OHLSSON, OLOF KARLSSON, AND THOMAS EDLUND*

Department of Microbiology, University of Umeå, S-901 87 Umeå, Sweden

Communicated by Viktor Mutt, February 29, 1988

ABSTRACT **The selective transcription of the Rat insulin 1 gene is mainly dependent on a beta-cell-specific enhancer element located in the 5′ flanking DNA. In analogy to many other viral and cellular enhancers, the insulin enhancer has been shown to be of a mosaic structure and the cis-acting elements of importance for the enhancer activity have been defined. Two short sequences are of crucial importance for the enhancer activity since mutation of either sequence leads to a decrease in activity (by a factor of ≈10), and the double mutant eliminates all enhancer activity. This study shows that these two major cis-acting elements interact with beta-cell-specific proteins. These two enhancer modules carry an 8-base-pair homology and compete with each other for protein binding, suggesting that they interact with the same protein, designated insulin enhancer binding factor 1 (IEF 1). Since mutation of these sequences eliminates the enhancer activity and protein binding, we propose that IEF 1 is the key regulator controlling the selective activity of the insulin gene enhancer.**

The mechanisms underlying cell-specific expression of mammalian genes are poorly understood, although a variety of tissue or cell-specific regulatory DNA elements such as enhancers (1–6) and promoters (1, 7–11) have been identified. The tissue specificity of such elements has in some systems been correlated to the presence of cell-specific nuclear proteins capable of interacting with various upstream promoter elements and enhancer sequences (12–18).

The Rat insulin 1 gene enhancer, which is located between positions −103 and −332 of the 5′ flanking DNA, plays a major role in controlling the cell-specific expression of the gene (1, 19). A recent systematic mutational analysis has revealed that the enhancer is composed of multiple short DNA elements, all of which contribute to the maximal activity of the enhancer (20). The two most affected block-replacement mutants located at positions −104 to −112 and −233 to −241 show a decrease in activity (by a factor of ≈10), and a double mutant eliminates all detectable enhancer activity (20).

We have previously shown that three distinct regions (E1, E2, and E3) within the enhancer are protected from DNase I cleavage in nuclear extracts prepared from insulin-producing cells (14). The protected region E1 was restricted to extracts prepared from insulin-producing cells, whereas E2 and E3 could also be detected in extracts from heterologous cells. We have now extended that analysis and focused our efforts on the recently defined two major determinants of enhancer activity located at positions −104 to −112 and −233 to −241 (20). By making use of the specific enhancer mutations discussed above, we can now conclusively show that these two short cis-acting elements act as binding sites for nuclear proteins. By using short synthetic DNA fragments spanning these sequences in combination with the comparatively more sensitive gel-mobility shift technique (21), we also show that the proteins interacting with these two sequences are present in two different insulin-producing cell lines but are absent in a variety of other cell types.

These two cis-acting DNA elements contain an 8-base-pair (bp) homology, GCCATCTG, and they compete with each other in the protein-binding assay, strongly suggesting that these major enhancer DNA elements interact with the same beta-cell-specific protein.

MATERIALS AND METHODS

Cell Lines and Preparation of Nuclear Extracts. Nuclear extracts were prepared as described (14) from the following cell lines: HIT M2.2.2, an insulin-secreting pancreatic beta-cell line derived from hamster (1); 5AHT2, an insulin-secreting subclone of the rat islet tumor cell line RIN-m 5F (22); BHK 21; and CHO. All cell lines were grown in tissue culture flasks prior to the preparation of extracts.

DNase Protection ("Footprint") Analysis. The preparation of DNA fragments and footprint reactions were performed as described (14). The DNA fragments used represent wild-type or mutant alleles of the 5′ flanking region of the Rat insulin 1 gene. The 5′ flanking mutants used are described elsewhere (20).

Gel-Mobility Shift Analysis. Oligonucleotides

```
5′ GATCCGCCATCTGCCA 3′        A
   3′ GCGGTAGACGGTCTAG 5′

5′ GATCCGGCCATCTGGCA 3′       B
   3′ GCCGGTAGACCGTCTAG 5′
```

and

```
5′ GATCCGCCAATCTGCCA 3′       C
   3′ GCGGTTAGACGGTCTAG 5′
```

were custom made by SYN-TEK AB (Umeå, Sweden) and designed to give 5′ protruding ends following hybridization. The hybridized oligonucleotides were made fully double-stranded and simultaneously were labeled in a fill-in reaction using DNA polymerase I Klenow fragment and [α-^{32}P]dATP; they were subsequently purified from polyacrylamide gels. The standard binding reaction contained the following components in a final volume of 8 μl: 25 mM Hepes (pH 7.9), 150 mM KCl, 10% glycerol, 5 mM dithiothreitol, 5000 cpm (1–3 fmol) of end-labeled, double-stranded oligonucleotide, 300 ng of poly[d(I-C)], 300 ng of poly[d(A-T)], 250 fmol of single-stranded heterologous oligonucleotide, and 3 μg of crude nuclear extract. The specific end-labeled synthetic DNA fragment was added last and the mixture was incubated for 15–20 min at 25°C. Following binding, the mixture was loaded directly onto a 4% polyacrylamide gel (acrylamide:bisacryl-

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

Abbreviation: IEF 1, insulin enhancer binding factor 1.
*To whom reprint requests should be addressed.

amide ratio, 40:0.8) containing 40 mM Tris/glycine (pH 8.5) (21). The gel was electrophoresed for 50 min at 25 mA with 40 mM Tris/glycine (pH 8.5) as running buffer, dried, and subjected to autoradiography using an intensifying screen at −70°C.

In the competition experiments the conditions were as described above except that unlabeled specific or nonspecific competitor fragment was included in the mixture prior to the addition of extract. A 25-fold excess of the competitor was used.

RESULTS

The Two Major Determinants of Enhancer Activity Interact with Nuclear Proteins. Since the DNA fragment used in our previous footprint analysis was labeled at position −103 it was difficult to detect protection of sequences −104 to −112 (14). We have therefore now used a DNA fragment labeled further downstream, at position −85, and can now, in addition to the previous noted protections (E1, E2, and E3) (14), observe protection of sequences −103 to −122 (E4) and −224 to −257 (E5) (Fig. 1*A*). To confirm that protections E4 and E5 were specifically due to interactions of factors with the two short mutationally sensitive sequences −104 to −112 and −233 to −241, respectively, the footprint patterns of DNA fragments carrying mutations in these sequences were analyzed. Mutation of sequences at positions −104 to −112 clearly eliminates the E4 footprint (Fig. 1*B*). The E5 footprint is analogously abolished by mutation of sequences at positions −233 to −241 (Fig. 1*C*). Interestingly, the E5 mutant does not affect the previously defined closely located E2 footprint (14), demonstrating that the juxtaposed protected regions E2 and E5 constitute two distinct protein-binding domains.

No protections except for the previously defined E2 and E3 regions (14) were detected in extracts prepared from the heterologous CHO and BHK 21 cell lines (data not shown).

Beta-Cell-Specific Proteins Bind to the Two Major Cis-Acting Determinants of Insulin Enhancer Activity. To confirm the cellular distribution of the interacting factors we used the comparatively more sensitive gel-mobility technique (21). Complementary synthetic oligonucleotides carrying the insulin enhancer sequences −103 to −113 and flanking non-insulin sequences (see *Materials and Methods*) were hybridized, end-labeled, and incubated with 3 μg of nuclear extracts prepared from two insulin-producing cell lines, 5AHT2 and HIT M2.2.2, and two heterologous cell lines, CHO and BHK 21. These mixtures were subsequently analyzed by the gel-mobility shift technique (21). As shown in Fig. 2*A*, lanes 1 and 2, this synthetic DNA fragment (fragment A; *Materials and Methods*) was specifically retarded only in extracts prepared from the two insulin-producing cells. The same result was obtained by using a synthetic fragment (fragment B) carrying the insulin sequences −229 to −239 (Fig. 2*B*, lanes 1 and 2). In addition, the complexes obtained with these two different synthetic DNA fragments appeared at an identical position in the gel. No complexes were observed with any of these short fragments using the heterologous extracts (Fig. 2*A* and *B*, lanes 3 and 4). Both of these complexes were visible with only 1 μg of HIT M2.2.2 and 5AHT2 extracts, whereas no complexes were observed even with 10 μg of heterologous extracts (data not shown).

The lack of complex formation in the heterologous extracts could be explained by the presence of inhibitory substances in these extracts. To eliminate this possibility, mixing experiments with HIT M2.2.2 extract and heterologous extracts were carried out and, as shown in Fig. 2*C*, lanes 1–3, no inhibition of complex formation was observed.

As a control for the overall activity of all of these extracts we used a mutated form of the synthetic DNA fragment

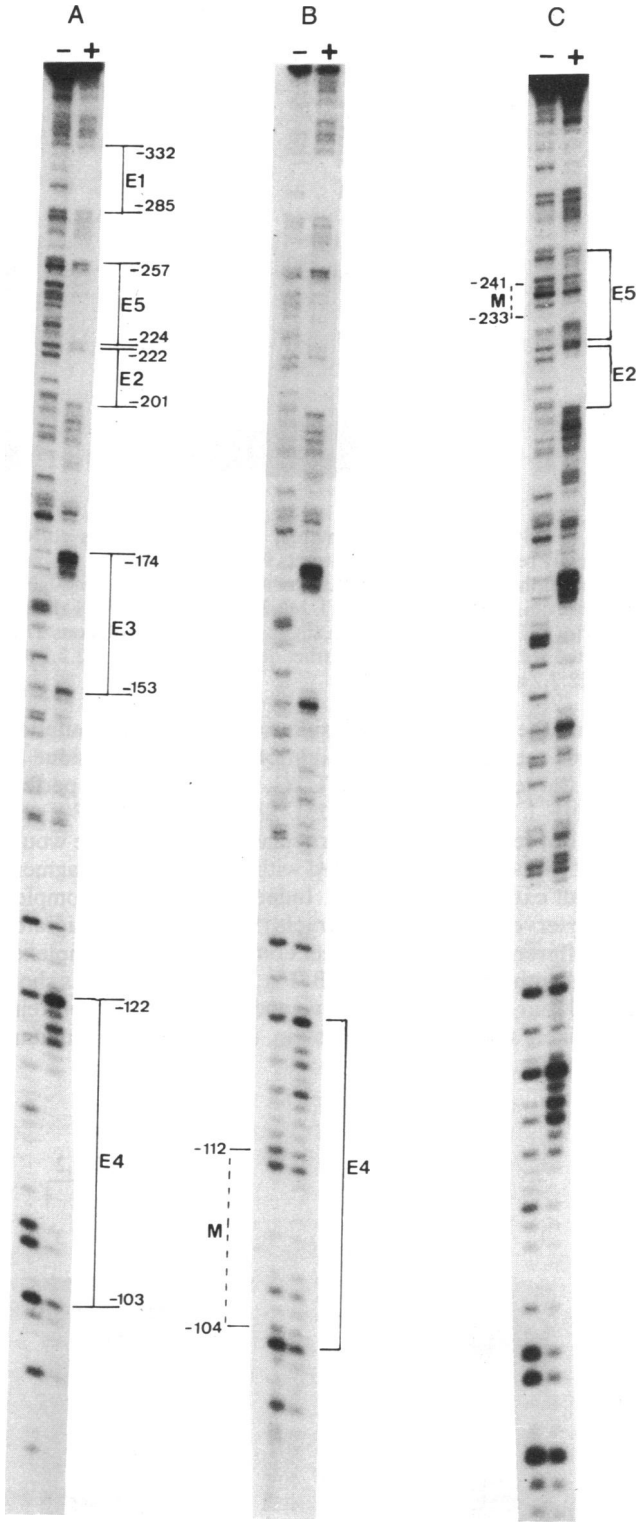

FIG. 1. Nuclear proteins from HIT M2.2.2 cells interact with the two major determinants of insulin enhancer activity. Wild-type and mutated forms of a 260-bp fragment of the insulin 5′ flanking DNA (−85 to −345) were subjected to footprint analysis by incubation with 150 μg of nuclear extract prepared from HIT M2.2.2 cells. Mutated sequences are indicated by an M and dotted bars. (*A*) Wild-type insulin fragment incubated with (+) or without (−) nuclear extract prior to DNase I treatment. (*B*) Same as *A* but with a fragment in which the sequence between positions −104 and −112 is mutated. (*C*) Same as *A* but with a fragment in which the sequence between positions −233 and −241 is mutated.

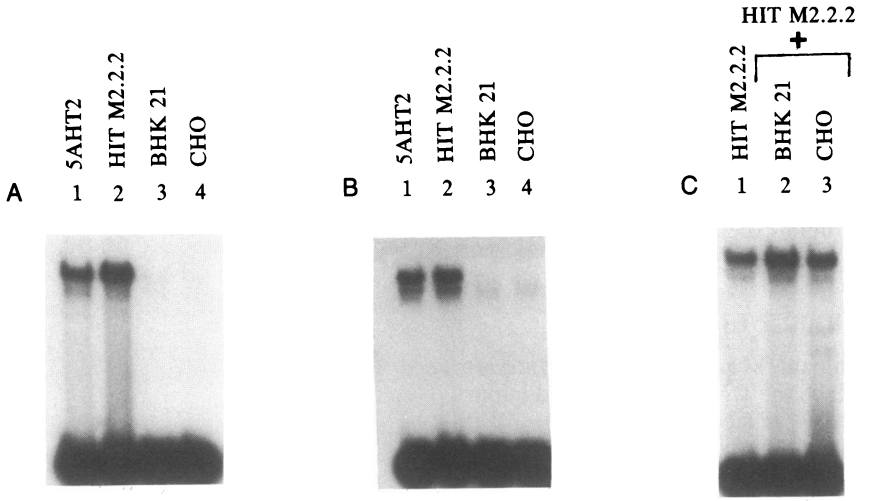

FIG. 2. Synthetic DNA fragments, carrying sequences corresponding to the two major determinants of insulin enhancer activity, are specifically retarded in nuclear extracts from beta cells. The gel-mobility shift assays were carried out using nuclear extracts from different cells. (*A*) Electrophoretic mobility shift assay using fragment A. End-labeled fragment (5000 cpm, 1–3 fmol) was incubated with 3 μg of nuclear extract derived from 5AHT2 (lane 1), HIT M2.2.2 (lane 2), BHK 21 (lane 3), and CHO cells (lane 4) in the presence of 300 ng of poly[d(I-C)], 300 ng of poly[d(A-T)], and 250 fmol of a single-stranded heterologous 27-bp oligonucleotide. (*B*) Same as *A* but with fragment B. (*C*) Fragment A was incubated with 2 μg of nuclear extract derived from HIT M2.2.2 (lane 1), 2 μg each from HIT M2.2.2 and BHK 21 (lane 2), and 2 μg each from HIT M2.2.2 and CHO (lane 3).

carrying the −103 to −113 insulin sequences. This mutation involves the insertion of an additional adenosine residue to create the sequence GCCAATCTG, which is almost a perfect CTF/NF-1 half-binding site (23). Since the CTF/NF-1 protein is not restricted to any particular cell type (22) we would expect to see a complex formed with this fragment (fragment C) in all extracts used above. Indeed, an identical complex was observed with all four extracts (Fig. 3*A*), which migrated to a different position in the gel compared to the complexes formed with fragments A and B (Fig. 3*B*). These results show that all four extracts were active and that a different protein, most likely the CTF/NF-1 protein, is binding to this mutant sequence.

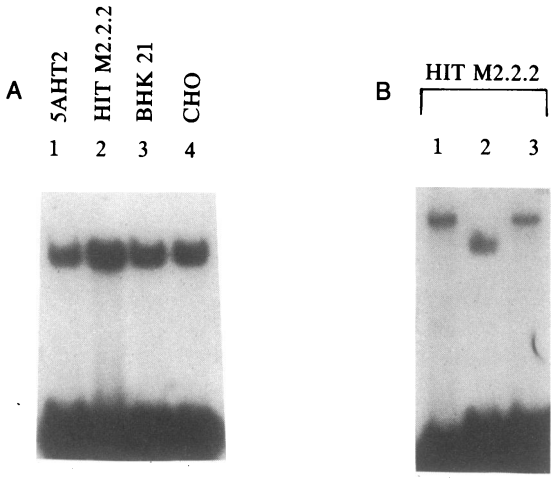

FIG. 3. Insertion of one nucleotide into the −103 to −113 enhancer sequence leads to the binding of a ubiquitous rather than a cell-specific protein. Fragment C, carrying an additional adenosine residue in the −103 to −113 enhancer sequence and resembling a CTF/NF-1-binding site, was analyzed in the gel-mobility shift assay by using extracts prepared from four different cells. See text for description of binding conditions. (*A*) Fragment C carrying a CTF/NF-1-binding site was incubated with 3 μg of 5AHT2 (lane 1), HIT M2.2.2 (lane 2), BHK 21 (lane 3), and CHO (lane 4) nuclear extracts. (*B*) Three micrograms of HIT M2.2.2 nuclear extract was incubated with the following synthetic fragments: fragment A (lane 1), fragment C (lane 2), and fragment B (lane 3).

The Two Major Determinants of Insulin Enhancer Activity Interact with the Same Cell-Specific Protein. To confirm the sequence specificity of the observed interactions, competition experiments were carried out by using the above described synthetic DNA fragments. Fragment A is specifically competitively inhibited with excess of the homologous fragment but not with fragment C (Fig. 4*A*, lanes 1–6). Analogously, fragment C is competitively inhibited only with the homologous fragment and not with fragment A (Fig. 4*C*, lanes 1–12). An identical result was obtained by using fragment B, except that the latter complex was also specifically competitively inhibited with fragment A (Fig. 4*B*, lanes 1–8). This cross-competition is not unexpected since the sequences GCCATCTCG at positions −105 to −112 and −231 to −238 are identical (ref. 16; *Materials and Methods*).

DISCUSSION

The cell-specific transcriptional activity mediated by the Rat insulin 1 5′ flanking DNA is, apart from the "TATA box," mainly dependent on two short homologous sequences, which are both located within the enhancer at positions −104 to −112 and −233 to −241 (20). Two other regions have been identified within the enhancer that contribute less, but significantly, to the activity of the enhancer. The maximal insulin enhancer activity is therefore dependent on the combined action of all of these elements (20).

We have earlier reported the identification of three different footprints within the insulin enhancer by using crude nuclear extracts prepared from insulin-producing cells (14). The protected region E1 was restricted to extracts prepared from insulin-producing cells, whereas E2 and E3 could also be detected in extracts prepared from heterologous cells. Interactions E1 and E2 correlate with the last two enhancer regions discussed above, which, when individually mutated, resulted in a decrease of enhancer activity by a factor of 3 (20). Interaction E3 is also correlated to a mutational sensitive region, although the effect of mutation of these sequences is even smaller, a factor of ≈2 (14, 20). This correlation, between elements important for the activity of the enhancer and DNA elements that interact with nuclear proteins from pancreatic beta-cells, indicates that there is a transcriptional regulatory function for these DNA-binding

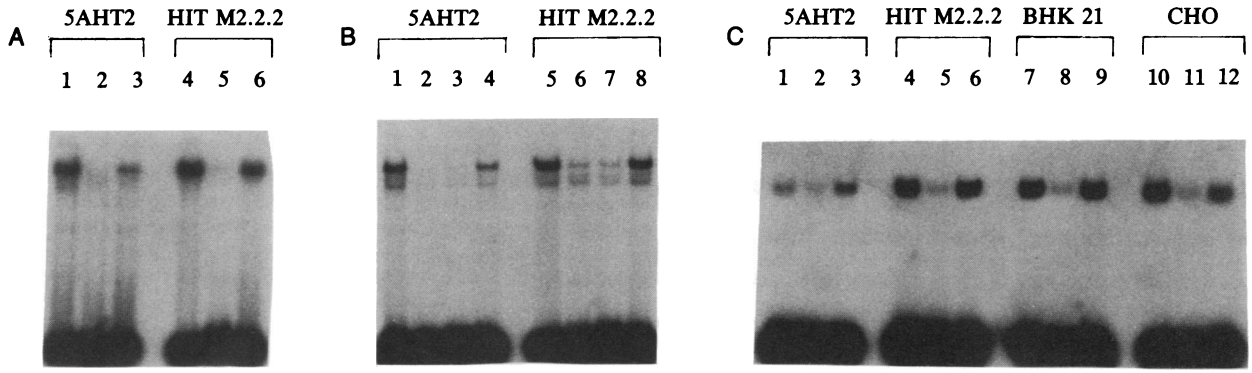

FIG. 4. The same cell-specific protein is binding to the two major determinants of enhancer activity. Fragments A–C underwent competitive inhibition against each other for the binding of nuclear proteins. See text for description of binding conditions. (*A*) Competition analysis of the complex formed with fragment A in 5AHT2 (lanes 1–3) and HIT M2.2.2 (lanes 4–6) extracts. Twenty-five femtomoles of the respective competitor fragment was included in the reaction mixture. Lanes 1 and 4, no competitor; lanes 2 and 5, the homologous fragment as competitor; lanes 3 and 6, fragment C as competitor. (*B*) Same as *A* using fragment B in 5AHT2 (lanes 1–4) and HIT M2.2.2 (lanes 5–8) nuclear extracts with the following competitors: no competitor (lanes 1 and 5), the homologous fragment (lanes 2 and 6), fragment A (lanes 3 and 7), and fragment C (lanes 4 and 8). (*C*) Same as *A* using fragment C in 5AHT2 (lanes 1–3), HIT M2.2.2 (lanes 4–6), BHK 21 (lanes 7–9), and CHO (lanes 11 and 12) nuclear extracts. Lanes 1, 4, 7, and 10, no competitor; lanes 2, 5, 8, and 11, the homologous fragment as competitor; lanes 3, 6, 9, and 12, fragment A.

proteins *in vivo* and that the beta-cell-specific nuclear protein(s), interacting with region E1, directly contributes to the cell-specific activity of the insulin enhancer. In this study, using footprint analysis of wild-type and mutant insulin enhancer DNA fragments, we have demonstrated the interaction of proteins with the recently defined key elements of insulin enhancer activity. These interactions were not detected in extracts prepared from heterologous non-insulin-producing cells, implying a cell-specific distribution of the interacting factors.

The results obtained by using the gel-mobility shift technique (21) showed that these factors are restricted to insulin-producing cells and that these two sequences competed with each other for binding to protein. As pointed out earlier (20), these two regions contain an 8-bp homology, GCCATCTG, strongly suggesting that the DNA sequences at positions -104 to -112 and -233 to -241 are interacting with the same protein, designated insulin enhancer binding factor 1 (IEF 1).

Since mutation of these sequences eliminates the enhancer activity in beta cells and the interaction with IEF 1, we propose that IEF 1 is the key regulator of the machinery controlling the cell-specific activity of the insulin gene enhancer.

The similarity between the described 8-bp insulin sequence GCCATCTG and some CTF/NF1-binding sites, GCCAAT (23), is striking; still, the introduction of an extra adenosine residue into the insulin sequence results in a loss of enhancer activity by a factor of ≈ 10 (20) and leads to the formation of a different non-cell-specific complex in the gel-mobility shift assay. These results demonstrate the stringent sequence requirements of these DNA–protein interactions and the functional divergence of the interacting proteins.

In this study we have not addressed the question: At what level is the cell-specific distribution of this protein(s) regulated? The control could be exerted at the transcriptional level, at the translational level, or, as recently suggested for two factors (NF-A1 and NF-kB), interacting with immunoglobulin gene enhancers (12, 24, 25), at the level of post-translational modification. The isolation of IEF 1 and the corresponding gene should help to answer this question.

We thank Professor William Rutter and Dr. Michael Walker for valuable discussions. This work was supported by Swedish National Research Council Grant 4875-107, Swedish Medical Research Council Grant 07149, and National Swedish Board for Technical Development Grant 84-05425P.

1. Edlund, T., Walker, M. D., Barr, P. J. & Rutter, W. J. (1985) *Science* **230**, 912–916.
2. Banjeri, J., Olson, L. & Schaffner, W. (1983) *Cell* **33**, 729–740.
3. Gillies, S. D., Morrison, S. L., Oi, V. T. & Tonegawa, S. (1983) *Cell* **33**, 717–728.
4. Neuberger, M. S. (1983) *EMBO J.* **2**, 1373–1378.
5. Queen, C. & Baltimore, D. (1983) *Cell* **33**, 741–748.
6. Queen, C. & Stafford, J. (1984) *Mol. Cell. Biol.* **4**, 1042–1049.
7. Mason, J. O., Williams, G. T. & Neuberger, M. S. (1985) *Cell* **41**, 479–487.
8. Falkner, F. G. & Zachau, H. G. (1984) *Nature (London)* **310**, 71–74.
9. Picard, D. & Schaffner, W. (1985) *EMBO J.* **4**, 2831–2838.
10. Wright, S., Rosenthal, A., Flavell, R. & Grosveld, F. (1984) *Cell* **38**, 265–273.
11. Bodner, M. & Karin, M. (1987) *Cell* **50**, 267–275.
12. Sen, R. & Baltimore, D. (1986) *Cell* **47**, 921–928.
13. Staudt, L. M., Singh, H., Sen, R., Wirth, T., Sharp, P. A. & Baltimore, D. (1986) *Nature (London)* **323**, 640–643.
14. Ohlsson, H. & Edlund, T. (1986) *Cell* **45**, 35–44.
15. Emerson, B. M., Lewis, C. D. & Felsenfeld, G. (1985) *Cell* **41**, 21–30.
16. Ephrussi, A., Church, G. N., Tonegawa, S. & Gilbert, W. (1985) *Science* **227**, 134–139.
17. Landolfi, N. F., Capra, J. D. & Tucker, P. W. (1986) *Nature (London)* **323**, 548–551.
18. Cereghini, S., Raymondjean, M., Carranca, A. G., Herbomel, P. & Yaniv, M. (1987) *Cell* **50**, 627–638.
19. Walker, M. D., Edlund, T., Boulet, A. M. & Rutter, W. J. (1983) *Nature (London)* **306**, 557–561.
20. Karlsson, O., Edlund, T., Barnett, J., Rutter, W. J. & Walker, M. D. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 8819–8823.
21. Schneider, R., Gander, I., Müller, U., Mertz, R. & Winnacker, E. L. (1986) *Nucleic Acids Res.* **14**, 1303–1317.
22. Bhathena, S. J., Oie, H. K., Gazdar, A. F., Voyles, N. R., Wilkins, S. D. & Recant, L. (1982) *Diabetes* **31**, 521–531.
23. Jones, K. A., Kadonaga, J. T., Rosenfeld, P. J., Kelly, T. J. & Tjian, R. (1987) *Cell* **48**, 79–89.
24. Gerster, T., Matthias, P., Thali, M., Jiricny, J. & Schaffner, W. (1987) *EMBO J.* **6**, 1323–1330.
25. Atchinson, M. L. & Perry, R. P. (1987) *Cell* **48**, 121–128.