# A cDNA from a mouse pancreatic β cell encoding a putative transcription factor of the insulin gene

Michael D.Walker*, Cheol Won Park, Ada Rosen and Ami Aronheim
Department of Biochemistry, Weizmann Institute of Science, 76100 Rehovot, Israel

Received December 5, 1989; Revised and Accepted January 25, 1990 EMBL accession no. X17500

## ABSTRACT

**Cell specific expression of the insulin gene is achieved through transcriptional mechanisms operating on multiple DNA sequence elements located in the 5′ flanking region of the gene. Of particular importance in the rat insulin I gene are two closely similar 9 bp sequences (IEB1 and IEB2): mutation of either of these leads to 5 – 10 fold reduction in transcriptional activity. We have screened an expression cDNA library derived from mouse pancreatic endocrine β cells with a radioactive DNA probe containing multiple copies of the IEB1 sequence. A cDNA clone (A1) isolated by this procedure encodes a protein which shows efficient binding to the IEB1 probe, but much weaker binding to either an unrelated DNA probe or to a probe bearing a single base pair insertion within the recognition sequence. DNA sequence analysis indicates a protein belonging to the helix-loop-helix family of DNA-binding proteins. The ability of the protein encoded by clone A1 to recognize a number of wild type and mutant DNA sequences correlates closely with the ability of each sequence element to support transcription in vivo in the context of the insulin 5′ flanking DNA. We conclude that the isolated cDNA may encode a transcription factor that participates in control of insulin gene expression.**

## INTRODUCTION

In mammals, biosynthesis of the protein hormone insulin is restricted to the β cells of the endocrine pancreas. This involves transcriptional control operating via sequences located in the 5′ flanking region of the gene (1,2). The transcription regulatory region of the rat insulin I gene has been studied in some detail: it contains a β cell specific enhancer and a β cell specific promoter (3). These DNA sequences act as positive control elements—in β cells they activate transcription, whereas in non-β cells they show no effect. Other experiments suggest that negative control plays a role in suppressing transcription in non-β cells (4,5).

Systematic mutational analysis has shown that within the 5′ flanking region, the most important positive elements of the cell specific response are two closely related 9 bp sequences, IEB1 (GCCATCTGC) and IEB2 (GCCATCTGG), located at −108 and −233 respectively in the rat insulin I gene (6). These sequences are capable of activating transcription in cell specific fashion when linked to a heterologous promoter (7,8). The simplest explanation of these results is that transcription factors that bind selectively to these sequences are present specifically in β cells. Indeed Ohlsson et al. (9) observed that nuclear proteins found selectively in β cells can interact with these sequences. On the other hand, Moss et al. (10) were able to detect binding activity to these sequences in whole cell extracts of both β and non-β origin. The IEB1 and IEB2 sequences are closely similar (10) to a series of elements termed 'E boxes' that were originally identified in immunoglobulin gene enhancers by in vivo footprinting techniques (11) and subsequently shown by transfection experiments to play an important role in expression (12).

To better understand the structure and function of the protein(s) which interact with the IEB1 and IEB2 sequences, we attempted to clone the cDNA(s) that encode them. The procedure was based on the recognition by recombinant phage protein of a radioactive DNA probe corresponding in sequence to the binding site of the desired protein (13,14). Two clones were isolated from a mouse β cell cDNA library on the basis of their reproducible binding to an IEB1 sequence probe. One of these, clone A1, shows strong sequence similarity to a human lymphocyte cDNA clone (E47) encoding a DNA binding protein possibly involved in immunoglobulin gene expression (15). The in vitro DNA binding specificity of the protein encoded by clone A1 is consistent with a role in insulin gene transcription.

## MATERIALS AND METHODS

### Construction and screening of the cDNA library

Total RNA was isolated from βTC1 cells (an established mouse line derived from a β cell tumor—ref 16) using the guanidinium thiocyanate-CsCl procedure (17). From 5 μg of this material, a library was constructed in the vector λgt11 (18) using published procedures (19). The library was screened by a modification (14) of the procedure of Singh et al (13). The DNA probe was prepared by ligating a 180 bp DNA fragment containing 9 tandem repeats of the IEB1 sequence (Fig 1) and labelling by nick translation. Bacteriophage which responded positively were purified by 3 rounds of plaque purification. DNA was isolated from these phage and digested with EcoRI to determine the size of the cDNA insert, which was subsequently subcloned to plasmid

* To whom correspondence should be addressed

vectors. Manipulations involving recombinant DNA were performed by standard procedures (20).

### Expression of cDNA insert in bacteria

The cDNA insert was subcloned to the expression vector pATH2 (21). To attain the correct reading frame both insert and BamH I digested vector were treated with Klenow polymerase to fill in protruding ends prior to ligation and transformation of competent E. coli HB101. Bacteria bearing the insert in the correct orientation were cultured in M9 minimal medium in the presence of β-indoleacrylic acid as described (22) to elevate production of the fusion protein. Bacteria were harvested and sonicated in the presence of sample buffer containing sodium dodecyl sulphate (SDS) (23) to solubilize proteins. The supernatant solution was then fractionated by SDS polyacrylamide gel electrophoresis. Gels were either stained with Coomassie brilliant blue or electro-blotted on to nitrocellulose sheets as described (22).

### Western and South-western blot analysis

For antibody binding (Western analysis), nitrocellulose sheets containing electro-blotted proteins were incubated for 2 h at room temperature in buffer 1 (10 mM sodium phosphate pH7.4, 150 mM NaCl, 0.1% non-fat milk (Carnation), and 0.05% Tween 20). Subsequently the sheets were incubated for 2h in sealed plastic bags containing buffer 1 with a 1:1000 dilution of rabbit anti-trpE antiserum. Following 3 washes for 10 min each with buffer 1, 0.125 μCi/ml $I^{125}$ -labelled S. aureus protein A (Amersham) was added. After 90 min incubation, sheets were washed 3 times for 10 min with buffer 1, air dried and exposed to X-ray film.

For DNA binding (South-western analysis), nitrocellulose sheets were subjected to a denaturation-renaturation procedure. Denaturation was performed in buffer containing 50 mM Tris-HCl pH8.0, 7 M guanidine hydrochloride (US Biochemical practical grade), 50 mM dithiothreitol (DTT), 2 mM ethylenediamine tetra-acetate (EDTA) and 0.25% non-fat milk. Incubation was for 60 min at room temperature. Filters were transferred to renaturation buffer containing 50 mM Tris-HCl pH 8.0, 100 mM NaCl, 2 mM DTT, 2 mM EDTA, 0.1% Nonidet-P40 (NP-40), and 0.25% non-fat milk for at least 12 h at 4°. Subsequently, treated sheets were incubated for an additional 12 h with radiolabelled probe ($5-10\times10^5$ cpm/ml) in TNE (50 mM Tris-HCl pH 8.0, 50 mM NaCl, 1 mM EDTA, and 0.25% non-fat milk). Finally sheets were washed 3 times for 10 min at room temperature in TNE, air dried, and exposed to X-ray film.

### Northern and Southern blot analysis

RNA was prepared by the guanidinium thiocyanate-CsCl procedure (17) followed by fractionation by one passage on oligo-dT cellulose (24). Samples of 5 μg were electrophoresed on formaldehyde-agarose gels (25), and blotted overnight to nitrocellulose. Blots were incubated with radioactive probe derived from the 1 kb insert A1 (Amersham multiprime labelling system— specific activity $10^9$ cpm/μg) and washed as described (25). To verify that RNA of comparable quality and quantity was present in all lanes, blots were subsequently reprobed with radioactive DNA corresponding in sequence to 18S rRNA (26).

Chromosomal DNA was prepared using standard procedures (20), digested with appropriate restriction enzymes and fractionated on agarose gels. Following transfer to nitrocellulose, radioactive probe prepared as above from the insert of clone A1 was applied. Following hybridization, filters were washed twice for 15 min each at 50° in 0.015 M NaCl, 0.0015 M Na Citrate (25).

### Mobility shift assay (27)

E. coli bearing pATH plasmids were cultured as described above to induce production of fusion protein. Cell extracts were prepared by urea extraction (28). Binding reactions contained 25 mM Hepes (pH 7.5), 1 mM EDTA, 5 mM DTT, 10% glycerol, 150 mM NaCl, 0.5 mM phenylmethylsulphonyl fluoride (PMSF), 900 ng poly-d(AT), 900 ng poly-d(IC), 5,000−10,000 cpm of double stranded oligonucleotide, corresponding to the IEB1 sequence, labelled by 'fill-in' reaction with Klenow polymerase (specific activity 2,000 cpm/fmole) and bacterial extract containing 10 μg of protein in a final volume of 20 μl. To determine the DNA binding specificity of the fusion protein, binding reactions were performed in the presence or absence of unlabelled DNA sequences. For convenience, these unlabelled sequences consisted of cloned DNA fragments (89 bp long) containing 2 copies of the test sequence inserted into the BamH I and Bgl II sites of the vector pOK2 (6). After 25 min at room temperature, the reaction mixtures were applied in the presence of 10% glycerol, 0.2% macerated agarose, and 0.01% bromophenol blue to 4% polyacrylamide gels containing 40 mM Tris-HCl, 12.5 mM sodium acetate, 3.3 mM EDTA (pH 7.9). Following electrophoresis for 2 h at 25 mA, gels were dried and exposed to X-ray film. Autoradiograms were subsequently quantitated by densitometry. The efficiency of competition was determined by comparison of competition efficiency with the homologous IEB1 sequence.

### Cell culture and transfections

HIT T15 M.2.2.2 cells and BHK cells were cultured as described previously (3,29). Cells were transformed by the calcium phosphate procedure (30) and the activity of CAT reporter plasmids was determined by normalization of CAT activity with the β-galactosidase activity produced by the pRSV-βgal internal control plasmid (3). The CAT reporter plasmids have been described previously (6,10). They are based on expression controlled by the intact flanking DNA (−345 to −1). Mutant sequences contain alterations in the IEB1 region only (−104 to −112).

### DNA sequencing

DNA sequencing was performed by the dideoxy procedure (31). Templates were double stranded plasmids containing portions of

| Designation | Sequence |
|---|---|
| IEB1 | g a t c c G C C A T C T G C c a |
| IEB2 | g a t c cgG C C A T C T G G c a |
| μE1 | g a t c c G C C A T C T T G c a |
| μE2 | g a t c c G C C A G C T G C c a |
| CCAAT | g a t c c G C C AAT C T G C c a |

**Figure 1.** Oligonucleotides used in this study. In each case only one strand is shown. Oligonucleotides corresponding to the other strand were designed to leave a GATC protruding end in all cases to facilitate ligation of DNA to plasmids. Upper case represents native DNA sequences and lower case represents sequences corresponding to restriction enzyme recognition sites. Underlined letters indicate an alteration in sequence as compared to the insulin IEB1/2 sequences.

the A1 insert cloned into the vector Bluescript M13+ (Stratagene). Sequencing reagents and procedures were from the 'Sequenase' kit (US Biochemicals). Both strands of the A1 insert were completely sequenced.

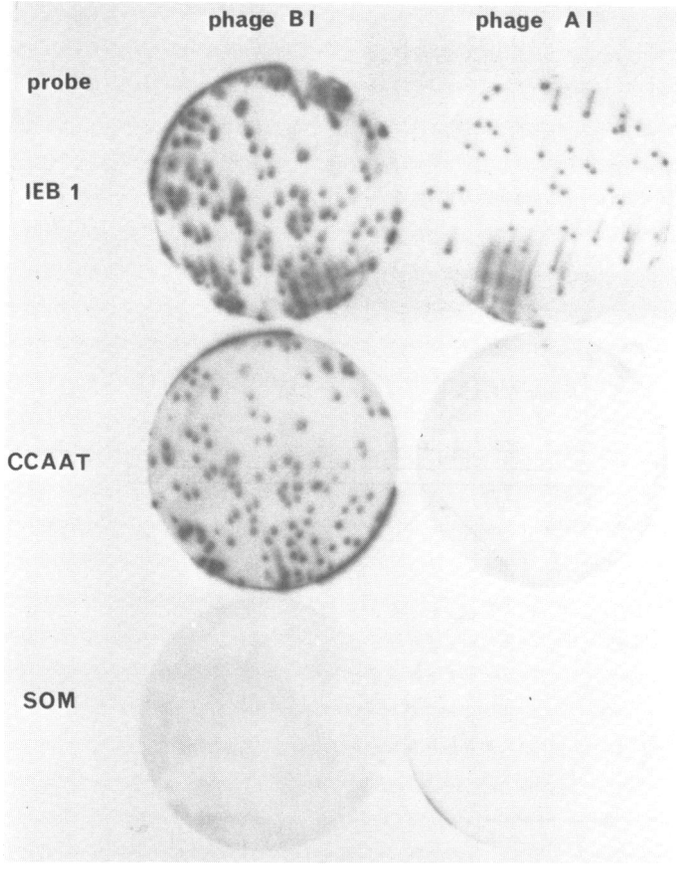


**Figure 2.** Autoradiogram of nitrocellulose filter lifts of plaque purified phage A1 and B1 derived from the βTC1 library. Triplicate filters were probed with radioactive DNA corresponding to the sequences of IEB1, CCAAT and an unrelated 124 bp DNA sequence (SOM) derived from the rat somatostatin gene (nucleotides −71 to +53, ref. 33).

## RESULTS

### Plaque screens

Screening of the library was performed using a radioactive probe prepared as described (14) from a cloned 180 bp DNA fragment containing 9 copies of the IEB1 sequence (Fig 1) in tandem. In a screen of 600,000 recombinant bacteriophages from the βTC1 library, we identified two clones (A1 and B1) which produced strong, reproducible signals with the above probe. These bacteriophage were plaque purified and characterized with respect to DNA binding specificity using 3 probes (see Fig 1 for details of the sequences): 1) The wild type IEB1 sequence used in the original screening. 2) A single nucleotide insertion to produce a 'CCAAT' sequence (32). 3) An unrelated 124 bp sequence derived from the rat somatostatin gene (33).

Although both clones interacted strongly with the original probe, clone B1 responded equally well with the point mutant, whereas clone A1 responded much less strongly (Fig 2). Neither phage produced a signal with the unrelated sequence. Since the CCAAT mutant sequence has previously been shown to have dramatically reduced transcriptional activity in vivo (6), in subsequent experiments we focussed attention on clone A1.

### Analysis of the protein encoded by A1

To characterize the cDNA present in this phage, the 1.0 kb EcoRI insert of clone A1 was subcloned to the bacterial expression vector pATH2. Bacteria bearing this plasmid, termed pATH-A1, showed high levels of production of a protein of molecular weight 67 kDa, absent in extracts of control pATH2-bearing bacteria (Fig 3a). The observed molecular weight is consistent with that of a fusion protein between trpE (37 kDa) and the protein coding potential of the 1 kb A1 insert. The fusion protein shows the expected binding to antibodies directed against trpE protein (Fig 3b). Furthermore a radioactive DNA probe corresponding to the IEB1 sequence was recognized by the protein (Fig 3c). This DNA binding activity was absent in extracts from bacteria bearing control pATH plasmid. A significantly weaker DNA binding activity of the fusion protein towards a mutant probe CCAAT (not shown) confirmed the binding specificity observed with bacteriophage plaques.

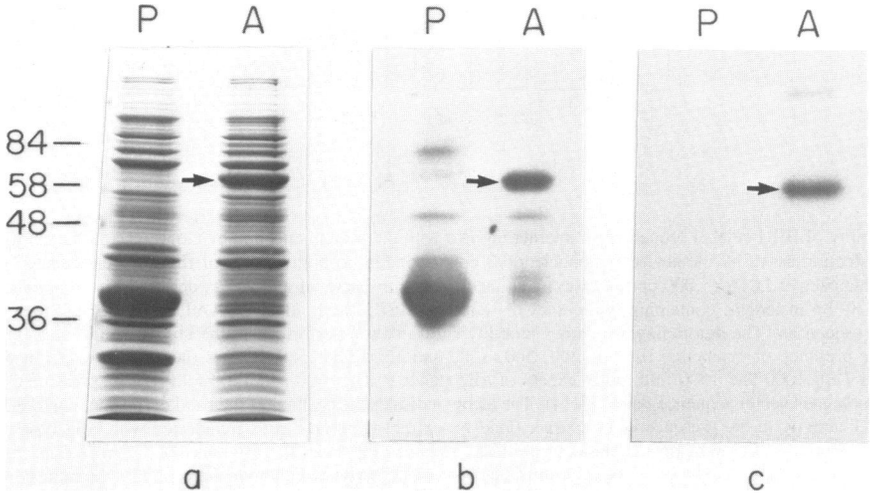


**Figure 3.** Analysis of bacterial fusion proteins. Bacterial extracts derived from pATH2 (lanes P) and pATH-A1 (lanes A) were fractionated on 10% SDS-polyacrylamide gels. Triplicate gels were treated either by staining with Coomassie brilliant blue (Panel a) or electro-blotting to nitrocellulose and probing with antibodies to trpE followed by $I^{125}$-labelled S. aureus protein A (Panel b) or probing with labelled IEB1 DNA (Panel c). Arrows mark the position of migration of the 67 kDa fusion protein.

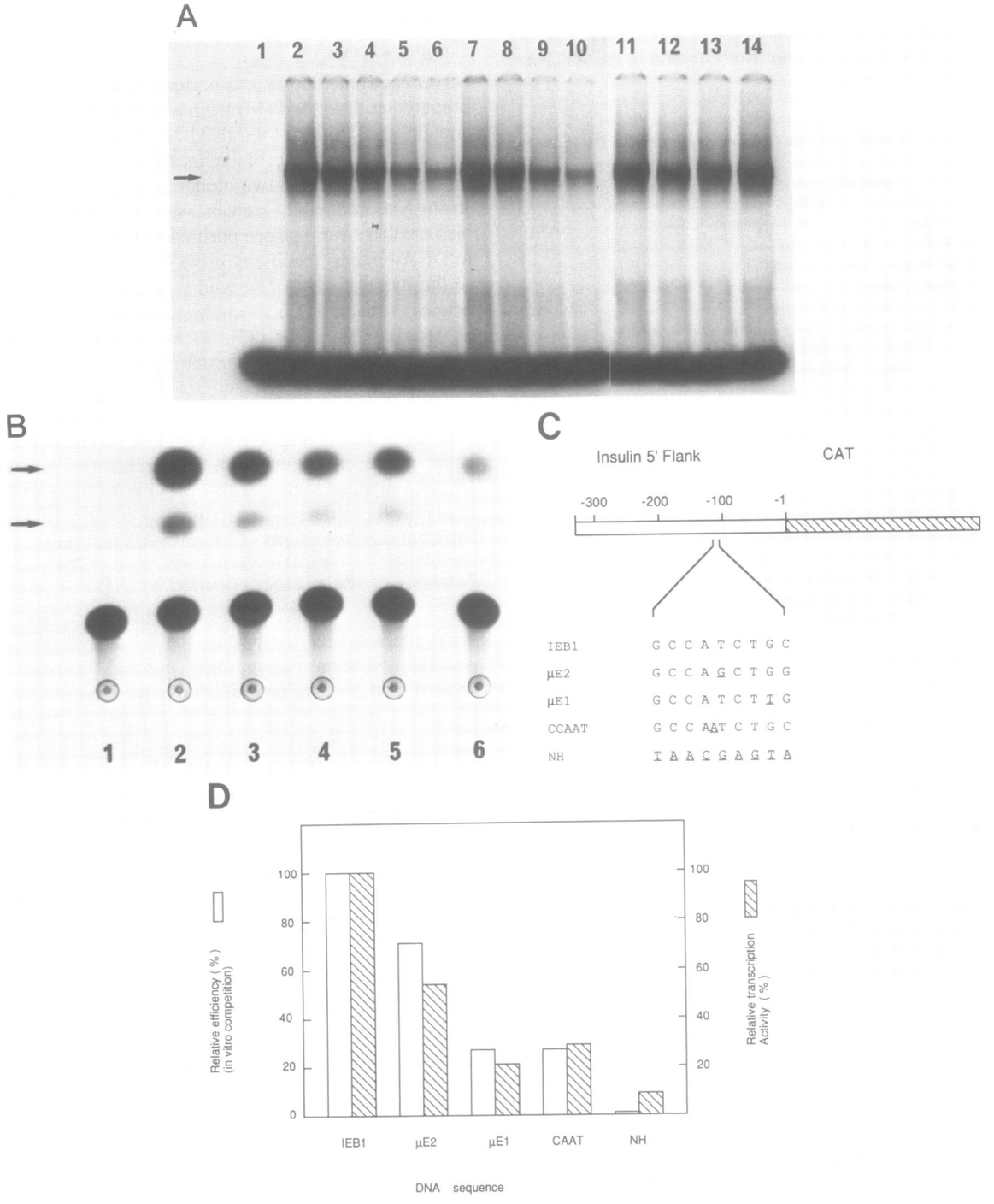


**Figure 4.** Comparison of the ability of IEB1-related sequences to interact in vitro with the protein encoded by clone A1 and the ability of these sequences to support transcription in an in vivo transfection assay. A. Annealed oligonucleotides corresponding to both strands of IEB1 were incubated in the presence of $\alpha^{32}$P dATP and Klenow polymerase. The radioactive DNA (5,000 cpm / assay) was incubated with bacterial extract, as described in Materials and Methods, in the presence or absence of unlabelled DNA 89 bp in length, containing two copies of sequences IEB1, μE1, μE2, CCAAT or the corresponding fragment from vector pOK2 (6) containing no IEB1 related sequences. The autoradiogram shows mobility shift analysis performed in the absence (lane 1) or presence (lane 2−14) of extract from pATH-A1 bacteria; in the presence of unlabelled IEB1 at 200, 500, 1000 and 1500 fold molar excess (lanes 3−6 resp); in the presence of 1000 and 1500 fold molar excess of μE1 (lanes 7,8); 1000 and 1500 fold molar excess of μE2 (lanes 9,10); 1000 and 1500 fold molar excess of CCAAT (lanes 11,12) and 1000 and 1500 fold excess of non-homologous vector sequence (lanes 13,14). The arrow indicates the position of the band corresponding to the DNA-fusion protein complex. B. Autoradiogram showing CAT activity in the presence of no extract (lane 1), extract from HIT cells transfected with wild type flank (lane 2) and extract from cells transfected with plasmids containing the following mutations in the IEB1 region: μE2 (lane 3), μE1 (lane 4), CCAAT (lane 5), and block mutant containing non-homologous (NH) sequences (lane 6). The arrows show the spots attributable to mono-acetylated chloramphenicol. C. Diagram showing structure of CAT plasmids used for comparisons of in vivo activity. D. Histogram comparing the above in vitro and in vivo activities. The ability of each sequence to compete for binding is compared to that of the homologous IEB1 sequence (open bars). Comparisons were performed using 250-fold molar excess of unlabelled DNA. Under these conditions the homologous IEB1 fragment caused a reduction of binding of 35%. In parallel, the effects of the same mutations on the activity of the intact 5′ flanking DNA was measured in a transfection assay (filled bars). All results shown represent the mean of 3 independent determinations.

### DNA binding specificity

Cell extracts were prepared from bacteria harboring the pATH-A1 recombinant plasmid and allowed to interact with an end-labelled DNA probe spanning the IEB1 binding site. The mixtures were fractionated on non-denaturing polyacrylamide gels and formation of protein–DNA complexes monitored by autoradiography of the dried gels. A strong shifted band of labelled DNA was observed when extract was prepared from bacteria harboring the recombinant plasmid (Fig 4a) but not from bacteria harboring the parental plasmid (not shown). Relative binding efficiency was estimated by performing the binding reactions in the presence of unlabelled DNA sequences, corresponding to wild type sequences from the insulin enhancer, or related sequences, including two 'E box' motifs μE1 and μE2 from the immunoglobulin heavy chain enhancer. The homologous IEB1 sequence (Fig 4a, lanes 3–6) and the IEB2 sequence (not shown) were the most efficient competitors. The μE2 sequence was somewhat less efficient while μE1 and CCAAT were poor competitors (Fig 4a, lanes 7–12). This correlates well with in vivo data (Fig 4b, refs 7,10) in which these sequences were tested for their ability to initiate transcription in the context of the intact insulin 5′ flanking DNA. The correlation between in vivo and in vitro activities is summarized in Fig 4d.

A higher molar excess of homologous competitor was required to substantially reduce signal strength in these assays than has been reported when cell extracts from mammalian cells were used in a similar test system (9,10). A likely explanation for this would be that the bacterially expressed protein has lower affinity for the cognate DNA sequence than the native protein, perhaps because of the prokaryotic trpE component or inappropriate post-translational modification.

### Analysis of structure and expression of the gene

The expression pattern of the gene was investigated by Northern blot analysis. Expression was found in a range of pancreatic and non-pancreatic cells (Fig 5). Highest levels were seen in the βTC1 cell line and in the spleen. A single hybridizing band of 3.4 kb was observed in all positive lines (Fig 5). The absence of a signal in pancreatic RNA (lane 6) is unexpected. However β cells represents only some 2% of total pancreas: most likely the amount of transcript is low in pancreatic exocrine cells (the predominant cell type of the pancreas). Genomic DNA from 2 established mouse cell lines, βTC1 and Ltk⁻ (mouse fibroblast) was prepared, digested with restriction endonucleases and analyzed by Southern blot procedures using a radioactive probe prepared from the 1.0 kb A1 cDNA. EcoR I and BamH I digestions resulted in single hybridizing bands of 19 kb and 13 kb respectively (Fig 6). Hind III digestion produced 2 bands of 15 and 7.5 kb. This pattern of hybridization is consistent with the idea that the genomic sequence is present in a single copy per haploid genome. Identical patterns were seen for the two cell types.

### Sequence of clone A1

The complete sequence of both strands of the 1.0 kb insert from clone A1 was determined by the dideoxy procedure following

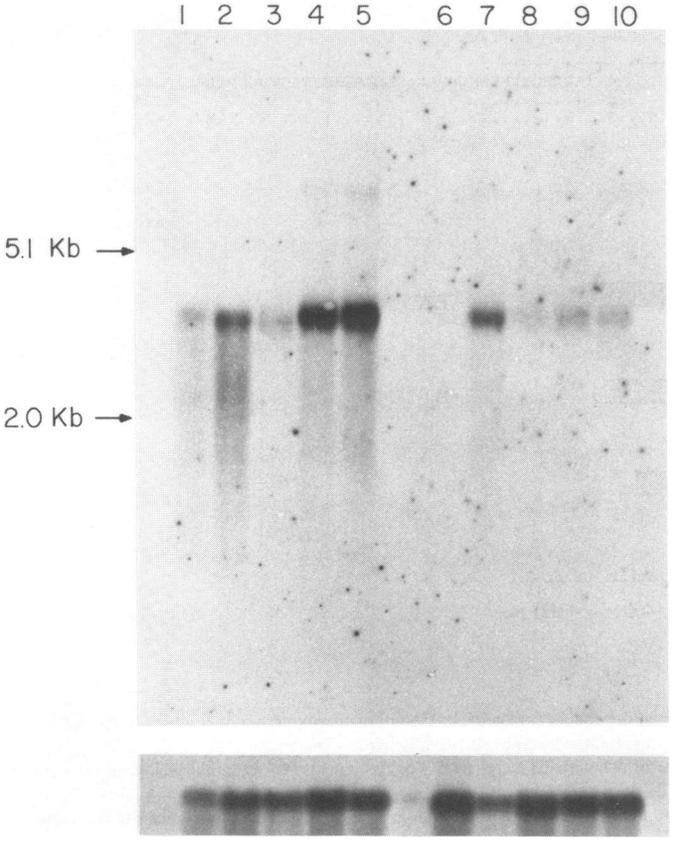


**Figure 5.** Analysis of RNA hybridizing to clone A1. RNA prepared as described was electrophoresed on formaldehyde-agarose gels and transferred to nitrocellulose. Blots were initially probed with radioactive DNA derived from clone A1 (main figure) and subsequently 18S rRNA (bottom inset). Lanes : 1—mouse T lymphoma cells (38); 2—mouse Ltk⁻ fibroblast cells; 3—hamster HIT cells; 4—mouse βTC1 cells; 5—mouse spleen; 6—mouse pancreas; 7—mouse lung; 8—mouse liver; 9—mouse brain; 10—mouse kidney. The migration of 28S rRNA (5.1 kb) and 18S rRNA (2.0 kb) is indicated by arrows.

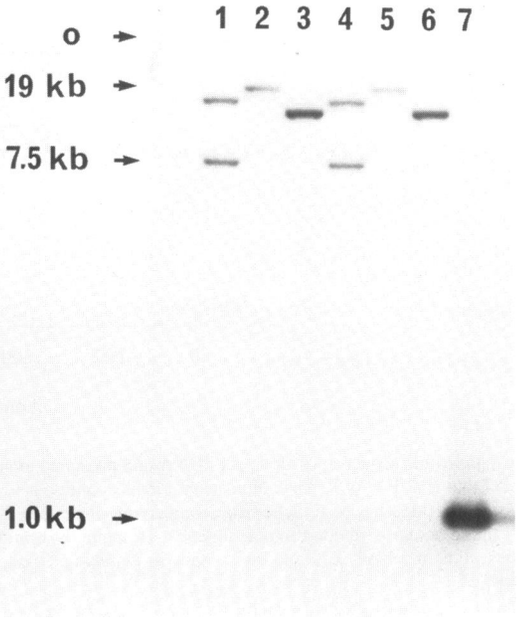


**Figure 6.** Analysis of genomic DNA corresponding to clone A1. Genomic DNA was isolated from βTC1 cells (lanes 1–3) or Ltk⁻ cells (lanes 4–6) and subjected to digestion with the restriction enzymes Hind III (lanes 1,4) EcoR I (lanes 2,5) and BamH I (lanes 3,6). Samples were electrophoresed on agarose gels, transferred to nitrocellulose filters and probed with a radioactively labelled clone A1 insert. Lane 7 contains 25 pg of the 1 kb insert of A1.

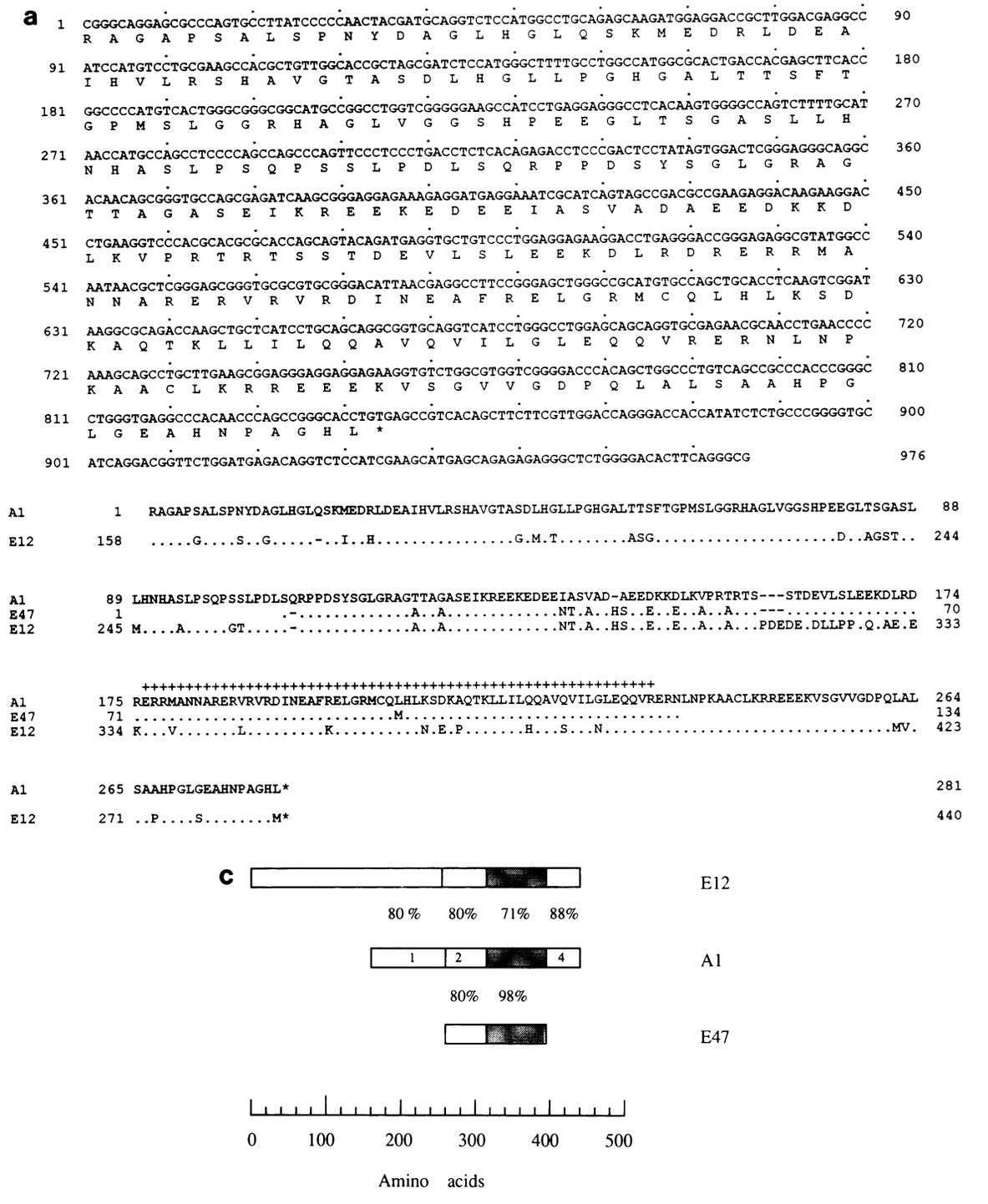


**Figure 7**. a. Nucleotide sequence of clone A1 cDNA and predicted amino acid sequence of the encoded protein. b. Comparison among amino acid sequence deduced from clone A1 and E12/E47 (15). Dots indicate an identical amino acid. Dashes indicate a gap introduced to permit alignment. The location of the putative heli-loop-helix domain is indicated by +. c. Schematic comparison of predicted protein sequence of A1 with E12 and E47. For the purpose of the comparison the sequence of A1 (281 amino acids) is divided into 4 domains. Domain 1—amino acids 1—104; 2—105—159; 3—160—232; 4—233—281. The extent of similarity between corresponding domains of A1, E12 and E47 is shown. Domain 3 (shaded) corresponds to the DNA binding domain of E12 and E47 and encompasses the helix-loop-helix domain (15).

subcloning of portions of the insert to Bluescript vectors (Fig 7a). The sequence shows an open reading frame of 281 amino acids with a potential initiation codon AUG (34) at amino acid 25. This could represent an authentic in vivo translation start site. On the other hand, the length of the RNA (3.4 kb) indicates that translation start sites further upstream are likely. The sequence of clone A1 shows strong similarity (Fig 7b) to a human cDNA clone (E47) which may encode a transcription factor involved in immunoglobulin gene transcription (15). The conservation is particularly strong (98% conservation at the amino acid level—Fig 7b,c) in a region possessing a novel DNA binding structure, the helix-loop-helix motif (15).

## DISCUSSION

Striking progress has been made in recent years in our understanding of the molecular basis for cell specific gene expression. In most cases, control is exerted at the transcriptional level via short DNA sequence elements located in the vicinity of the gene. It is generally believed that the importance of these sequences lies in their ability to interact with the DNA binding domain of specific transcription factors which are thereby positioned appropriately to permit activation of transcription of the nearby gene. Therefore a major focus of research in this area has become the isolation and characterization of these transcription factors and the genes which encode them, with a view to understanding how the factors function and how they in turn are controlled.

Mammalian insulin genes are an example of a gene family which exhibits virtually complete cell specificity in its expression pattern: high levels of mRNA are found in endocrine $\beta$ cells whereas essentially no insulin mRNA is found in other cell types. Although the insulin gene transcriptional control region is complex, a dominant component consists of the 9bp sequences (IEB1 and IEB2) which appear to be centrally involved in cell specific expression : upon mutation of either or both of these sequences, transcription from the insulin 5′ flank is substantially reduced (6). Furthermore, these sequences, in the absence of other sequences from the 5′ flank, are capable of activating a heterologous promoter in cell specific fashion (7,8). It therefore seems likely that the protein(s) interacting with IEB1 and IEB2 constitute an important component of the machinery which determines cell specific transcription of the insulin gene.

To complement biochemical approaches taken to identify and characterize such proteins (9,10), we undertook to clone the gene(s) encoding them using an expression screening approach. We have identified a cDNA clone that encodes a protein whose DNA binding specificity corresponds closely to that expected for a transcription factor that interacts with the IEB1 sequence. Analysis of the RNA of several mouse cell types indicates the presence of this or a closely related RNA in both pancreatic $\beta$ cells and non-pancreatic cells. Southern blot analysis reveals a pattern of bands consistent with a single copy gene.

Murre et al (15) used an approach similar to that described here to isolate 2 cDNAs (E12 and E47) of human lymphoid origin which specifically bind to an E box motif ($\kappa$E2 -GCCACCTGC). The sequence of these clones, and E47 in particular, is closely similar to that of the clone isolated in our study (Fig 7b,c). Given this high similarity and Southern blot analysis of mouse (Fig 6) and human (35) indicating a single copy per haploid genome, it seems likely that the mouse A1 cDNA and the human E47 cDNA represent transcripts produced from gene homologues. Whether the cDNAs themselves represent homologous transcripts is hard to determine at this point. In fact several related cDNAs have been identified in both mouse (our unpublished results) and human (15) cells. These cDNAs appear to contain similar 5′ regions but divergent 3′ regions. This complicates the elucidation of full length cDNA sequence (3.4 kb) corresponding to A1 and other clones since it is necessary to isolate individual clones containing essentially full length sequence rather than a series of shorter overlapping clones. Determination of the relationship between A1 and E47 cDNAs will ultimately require the full length sequence determination of both these and the related cDNAs of mouse and human.

How might the protein encoded by clone A1 be involved in $\beta$ cell-specific transcription? A simple possibility would be that the mRNA is selectively translated or post-transcriptionally modified. Alternatively the protein may be present in identical form in many cell types, but acquire ability to trans-activate in specific fashion as a consequence of differential accessibility to cognate DNA sequences (36) or by virtue of interaction with different, as yet unidentified protein(s) found only in $\beta$ cells. Such protein(s) may show intrinsic binding affinity to the IEB1 sequence or interact with the A1 protein but not with DNA. The ability of E12 and E47 proteins to form DNA-binding heterodimers between each other and with other helix-loop-helix proteins such as MyoD (37), an activator of the muscle phenotype, is consistent with these possibilities. In addition to clone A1 which we have characterized, we have isolated 3 additional cDNA inserts from the $\beta$TC1 library that are similar, but not identical to A1, as determined by restriction enzyme analysis (not shown). These may be formed by alternate splicing of the RNA transcribed from the gene as suggested for the E12 and E47 transcripts (15). The consequences at the protein level of such a pattern of expression of several non-identical transcripts are unclear. More detailed characterization of the transcripts and the encoded proteins will help to delineate the precise role of this interesting gene in insulin as well as immunoglobulin gene transcription.

## ACKNOWLEDGEMENTS

We would like to thank Dr. Yoav Citri and Dana Hevroni for advice and valuable assistance in construction of the cDNA library, Drs Doug Hanahan and Shimon Efrat for providing $\beta$TC1 cells, Dr Alpha Peled for T lymphoma cells, Drs Steve McKnight and Charles Vinson for instruction on expression screening, Drs Thomas Edlund and Olof Karlsson for gifts of plasmids and oligonucleotides and Drs Yoav Citri, Yosef Shaul, Thomas Edlund, Olof Karlsson, David Yaffe, Larry Moss and William Rutter for valuable discussions. The work was supported by grants to MDW by the NIH (GM38817), the Israel Academy for Sciences and Humanities and the Rockefeller-Weizmann Trust Fund. MDW was the recipient of a Career Development Award of the Juvenile Diabetes Foundation International.

## REFERENCES

1. Walker, M.D., Edlund, T., Boulet, A.M. and Rutter W.J. (1983) Nature, 306, 557–561.
2. Hanahan, D. (1985) Nature, 315, 115–122.
3. Edlund, T., Walker, M.D., Barr, P.J. and Rutter, W.J. (1985) Science, 230, 912–916.
4. Nir, U., Walker, M.D. and Rutter, W.J. (1986) Proc. Natl. Acad. Sci. (USA), 83, 3180–3184.
5. Laimins, L., Holmhren-Konig M. and Khoury G. (1986). Proc. Natl. Acad. Sci. (USA), 83, 3151–3155.
6. Karlsson, O., Edlund, T., Moss, J.B., Rutter, W.J. and Walker, M.D. (1987) Proc. Natl. Acad. Sci. (USA), 84, 8819–8823.
7. Karlsson, O., Walker, M.D., Rutter, W.J. and Edlund, T. (1989) Mol. Cell. Biol., 9, 823–827.
8. Whelan, J., Poon, D., Weil, P.A. and Stein, R. (1989) Mol. Cell. Biol., 9, 3252–3259.
9. Ohlsson, H., Karlsson, O. and Edlund, T. (1988) Proc. Natl. Acad. Sci. (USA), 85, 4228–4231.
10. Moss, L., Moss, J.B. and Rutter, W.J. (1988) Mol. Cell. Biol., 8, 2620–2627.
11. Ephrussi, A., Church, G.M., Tonegawa, S. and Gilbert, W. (1985) Science, 227, 134–140.
12. Lenardo, M., Pierce, J. and Baltimore, D. (1987) Science, 236, 1573–1577.
13. Singh, H., LeBowitz, J.H., Baldwin, A.S., Jr. and Sharp, P.A. (1988) Cell, 52, 415–423.
14. Vinson, C.R., LaMarco, K.L., Johnson, P.F., Landschulz, W.H. and

McKnight, S.L. (1988) Genes Dev., 2, 801–806.
15 Murre, C., McCaw, P.S. and Baltimore, D. (1989) Cell, 56, 777–783.
16. Efrat, S., Linde, S., Kofod, H., Spector, D., Delannoy, M., Grant, S., Hanahan, D. and Baekkeskov, S. (1988) Proc. Nat. Acad. Sci (USA), 85, 9037–9041.
17. Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) Biochemistry, 18, 5294–5299.
18 Young, R.A. and Davis, R.W. (1983) Science, 222, 778–782.
19 Gubler U. and Hoffman, B.J. (1985) Gene, 25, 263–269.
20. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory.
21. Dieckmann, C.L. and Tzagaloff, A. (1985) J. Biol. Chem., 260, 1513–1520.
22 Aronheim, A., Eshel, Y., Mosckovitz, R. and Gershoni, J.M. (1988) J. Biol. Chem., 263, 9933–9937.
23. Laemmli, U. (1970) Nature, 227, 680–685.
24. Aviv, H. and Leder, P. (1972) Proc. Natl. Acad. Sci. (USA), 69, 1408–1412.
25. Thomas, P. (1980) Proc. Natl. Acad. Sci (USA), 77, 5201–5205.
26. Bowman, L.H., Rabin, B. and Schlessinger, D. (1981) Nucl. Acids Res., 9, 4951–4966.
27. Fried, M. and Crothers, M.D. (1981) Nucl. Acids Res., 9, 6505–6525.
28. Kadonaga, J.T., Courey, A.J., Ladika, J.H. and Tjian, R. (1988) Science, 242, 1566–1570.
29. Santerre, R.F., Cook, R.A., Crisel, R.M.D. Sharp, J.D., Schmidt, R.J., Williams, D.C. and Wilson, C.P. (1981) Proc. Natl. Acad. Sci. (USA), 78, 4339–4343.
30. Wigler, M., Pellicer, A., Silverstein, S., Axel, R., Urlaub, G. and Chasin, L. (1979) Proc. Natl. Acad. Sci. (USA), 76, 1373–1376.
31. Sanger, F., Nicklen, S. and Coulsen, A. (1977) Proc. Natl. Acad. Sci (USA), 74, 5463–5467.
32. Chodosh, L.A., Baldwin, A.S., Carthew, R.W. and Sharp, P.A. (1988) Cell, 53, 11–24.
33. Tavianini, M.A., Hayes, T.E., Magazin, M.D., Minth, C.D. and Dixon, J.E. (1984) J. Biol. Chem., 259, 11798–11803.
34. Kozak, M. (1986) Cell, 44, 283–292.
35. Mellentin, J.D., Murre, C., Donlon, T.A., McCaw, P.S., Smith, S.D., Carroll, A.J., McDonald, M.E., Baltimore, D. and Cleary, M.L. (1989) Science, 246, 379–382.
36. Becker, P.B., Ruppert, S. and Schutz, G. (1987) Cell, 51, 435–443.
37. Murre, C., McCaw, P.S., Vaessin, H., Caudy, M., Jan, L.Y., Jan, Y.N., Cabrera, C.V., Buskin, J.N., Hauschka, S.D., Lassar, A.B., Weintraub, H. and Baltimore, D. (1989) Cell, 58, 537–544.
38. Peled, A., and Haran-Ghera, N. (1985) J. Exp. Med., 162, 1081–1086.