Vol. 170, No. 1, 1990 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
July 16, 1990 Pages 314-321

INSULIN ENHANCER BINDING PROTEIN HAS HELIX-LOOP-HELIX STRUCTURE

Yoshikazu Shibasaki*, Hiroshi Sakura,
Fumimaro Takaku, and Masato Kasuga

The Third Department of Internal Medicine, Faculty of Medicine,
University of Tokyo, 7-3-1 Hongo, Bunkyo, Tokyo, Japan 113

Received June 7, 1990

SUMMARY: Insulin gene expression is restricted to pancreatic B cells and the 5' flanking region is responsible for the tissue specificity. The GCCATCTG motif in this region of the rat insulin 1 gene functions as an enhancer for insulin transcription. A cDNA coding for a GCCATCTG motif-binding protein (IEBP1) was isolated from a rat pancreatic B cell tumor λgt11 library. The IEBP1 protein was found to be the rat counterpart of the immunoglobulin (Ig) enhancer binding protein E12/47 having a helix-loop-helix domain. This result indicates that the Ig gene and insulin gene employ the same (or a similar) binding protein as a part of their transcriptional apparatus.

In eukaryotic cells, the transcription of many genes is regulated through distinct cis-acting DNA elements which contain multiple binding sites for sequence specific DNA binding proteins. Several species of insulin gene have been cloned and transient expression experiments have shown that the 5'-flanking region can direct selective expression in a pancreatic endocrine derived cell line (HIT) (1,2). Using oligonucleotides covering the upstream region (-340 to +1) of the rat insulin 1 gene, systematic analysis has been carried out and four important regions have been found: (A) -32 to -23, (B) -112 to -104 (Nir box), (C) -241 to -233 (Far box), and (D) -309 to -302 (3). Whereas (A) is the well known TATA sequence, (B) and (C) include the same sequence motif GCCATCTG. This motif is conserved in rat 2 and human insulin genes. Similar sequences are found in Ig gene enhancers (E motif: μ heavy chain: μE1-μE4, κ light chain: κE1-κE3) (4-6). Synthetic oligonucleotide having this motif can enhance the transcription in a tissue specific manner when it is ligated to a

* To whom correspondence should be addressed.

0006-291X/90 $1.50

heterologous promotor (7). This motif appears to play a major role in the possitive control of B-cell-specific transcription of the insulin gene.

It has been suggested that sequence-specific DNA binding proteins interacting with these regions have functional importance for insulin gene expression. Five footprint regions in the 5' flanking region of rat insulin 1 gene have been found by using the HIT nuclear extract. Among them, E4 (-122 to -103) and E5 (-257 to -224) coincide with Nir and Far box, respectively (8,9). Nir probe (-93 to -134) and Far probe (-219 to -251) revealed several shifted bands by gel shift analysis. Since an oligonucleotide containing the Ig heavy chain μ enhancer E motif (μE) competed with these interactions, a similar factor to Ig enhancer μE2 binding factor is believed to be important for the insulin gene transcription (10).

Thus, it appears important to characterize these binding proteins. We report here a cDNA clone for a rat insulin enhancer binding protein and show that it has helix-loop-helix structure similar to those of Ig enhancer binding protein E12/47 (11).

MATERIALS AND METHODS

Oligonucleotides

```
GATCCGGCCATCTGGCA              RINS 1+2
    GCCGGTAGACCGTCTGA

TCGAGCCCCTCTCGCCATCTGCC        RINS 3+4
   TCGGGGAGAGCGGTAGACGGAGC

GATCCGGCCTAGTGGCA              MUT RINS 1+2
    GCCGGATCACCGTCTAG
```

were made with a DNA synthesizer. RINS 1+2 contained the Far box (-239 to -229) and RINS 3+4 contained the Nir box (-112 to -103) of rat insulin 1 gene. Central three bases of RINS 1+2 were changed in MUT RINS 1+2. Purified nucleotides were annealed in annealing buffer (20mM Tris-HCl pH 7.5, 10mM MgCl, 50mM NaCl and 1mM DTT), autoligated, and electrophoresed. Five repeats of oligonucleotides were then electroeluted, purified, nicktranslated with [^{32}P] dCTP, and finally repaired with T4 DNA ligase.

Poly (A)$^+$ containing RNA was purified from a rat insulinoma cell line (RINr). cDNA libraries were constructed in λgt11 with oligo (dT) or random hexanucleotide primers by using cDNA synthesis PLUS and cDNA cloning system-λgt11 (Amersham). The λgt11 library was screened with mixed five repeated RINS 1+2 and RINS 3+4 probes as described previously (13-15). Briefly, nitrocellulose filters were soaked in 10mM IPTG and air dried. LB plates containing phage

plaques that had been incubated for 3.5 hrs at 42^{0}C were overlaid with one filter and were incubated for 8 hrs. The filters were immersed for 2 hrs in 50mM Tris-HCl pH 7.5, 50mM NaCl, 1mM EDTA, 5μg/ml denatured salmon sperm DNA and 2x10^6 c.p.m./ml of [^{32}P] labeled probes. The filters were washed four times with the same buffer without probes for 30 min, air dried, and exposed to X-ray film overnight.

RESULTS AND DISCUSSION

By screening 1.2x10^6 recombinant phage plaques, two recombinants, IEBP-A and -B, were isolated. They did not cross-hybridize each other. The screening by this method often results in isolation of a clone for a nonspecific DNA binding protein (16). To examine the specificity, IEBP-A and -B were checked whether they bound to RINS 1+2 and MUT RINS 1+2. As shown in Fig. 1, IEBP-A

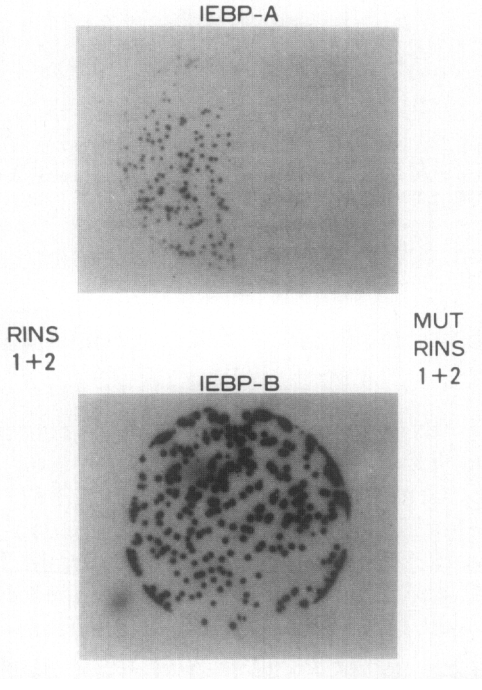

Fig. 1. Binding specificity of IEBP-A and -B to the insulin enhancer sequence. The rat RINr λgt11 library was screened with DNA probes containing five repeats of RINS 1+2 and RINS 3+4. IPTG induced proteins of purified phages were transferred to nitrocellulose filters. Two positive clones (IEBP-A, IEBP-B) were analyzed by Southwestern method with the wild type (RINS1+2) and mutated probes (MUT RINS 1+2). Each filter was cut in half and the fragments were reacted with each labeled oligomer.

bound only to RINS 1+2, but IEBP-B bound to both RINS 1+2 and MUT RINS 1+2. Therefore IEBP-B could not be a specific binding protein to the GCCATCTG motif.

The insert of IEBP-A containing 1kb of open reading frame (started at the junction between the IEBP-A and the β-galactosidase sequence) and the 0.6kb of 3' non-coding region was sequenced. Nucleotide sequence and deduced amino acid sequence of this protein (IEBP1) are shown in Fig. 2. There are several interesting features in the amino acid sequences of IEBP1. Repetition of leucine residues at every seventh position from amino acid 89 to 117, interrupted by the presence of a threonine, resembles the leucine

Fig. 2. Nucleotide sequence of the IEBP1 and the deduced amino acid sequence. DNA sequence of cloned cDNA was determined by the dideoxy method. Nucleotide position 1 is the start of the IEBP-A cDNA. Leucine residues in a leucine repeat are underlined. Open box indicates conserved 60 amino acids helix-loop-helix motif.

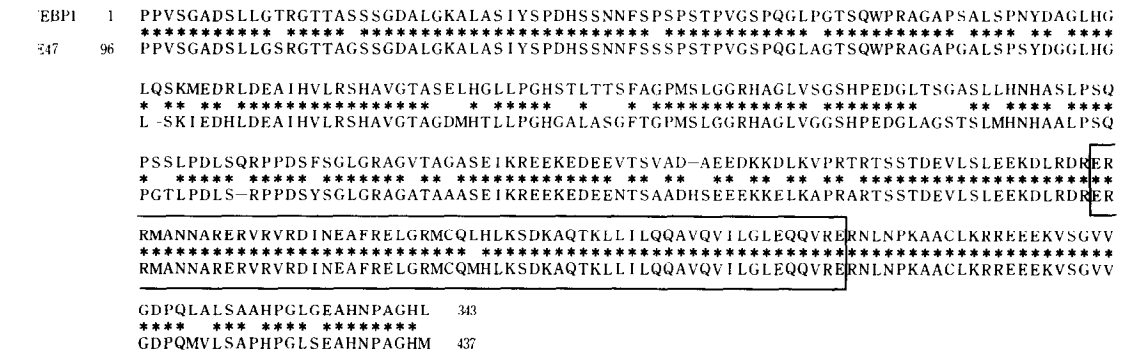

Fig. 3. Amino acid sequence relationship between IEBP1 and E47. Amino acid similarities between IEBP1 (a.a. 1 to 343, upper) and E47 (a.a. 96 to 437, lower) are shown. Asterisks indicate identical amino acid residues. Open box indicates helix-loop-helix.

zipper (17). The sequence of IEBP1 was closely related to that of E47 (11). As long as the amino acid sequence is available, 83% are identical. From amino acid 238 to 296 (58 out of 59 amino acids in this region are identical) (Fig. 3), the most probable secondary structure is two helices, first 12 and second 13 amino acids, which are connected by a loop like structure by analogy to E12/47. In the proposed helices, the hydrophobic residues (leucine, alanine, and phenylalanine) are conserved. Thus, IEBP1 seems to be a rat counterpart of human E47.

When poly $(A)^+$ RNA of RINr were hybridized with the insert of IEBP1 as a probe, 2.9 kb mRNA was found. On examination of the tissue distribution, a similar level of IEBP1 expression in rat hepatoma cell line H35 was observed (Fig. 4). The length was the same as that of RINr. Therefore, expression level of IEBP1 does not parallel that of insulin mRNA.

HLH proteins have been recently identified as DNA binding proteins which includes several important proteins (11,12). MyoD can differentiate fibroblasts into myotubes. The myc family of oncoproteins are also DNA binding proteins and known to regulate cell growth. It is interesting that HLH proteins include the genes that

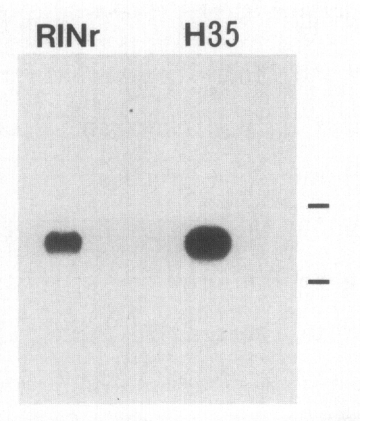

Fig. 4. Northern analysis. Poly (A)+ RNA were isolated from RINr and H35 cell lines. 20μg samples were electrophoresed and transferred to nylon membrane. Hybridization was carried out with the standard protocol by using insert of IEBP1 as a probe. Lines in autoradiogram indicate positions of 28S and 18S rRNA.

are related to differentiation (myoD, achaete-scute complex) and to the tissue specific genes (insulin, immunoglobulin).

The Nir oligonucleotide probe (-113 to -103) interacts with only insulin producing cell nuclear extract (HIT and 5AHT2) by gel mobility shift assay (9). It is possible that there are two different binding proteins that recognize the same DNA sequence. However, an alternative possibility is that some posttranslational modification is necessary for the DNA binding activity in vivo.

In the immunoglobulin enhancer, there are other (non E motif) protein binding elements, the octamer motif in heavy chain and the B motif in κ chain, which may determine the tissue restriction. Since IEBP1 mRNA is not pancreas specific, one can not explain the tissue specificity of insulin gene expression by IEBP1 and the GCCATCTG motif alone. However, it is suggested that HLH proteins are divided into three classes: class A (E12/47, daughterless), class B (myoD, achaete-scute complex), class C (myc) (12). Class A HLH proteins are ubiquitous components of a regulatory system that acts through enhancer sequence. They can form homo- or hetero-dimer to bind

DNA. For example, ubiquitous factor E12 can form hetero-dimer with muscle specific myoD and binds to the enhancer of muscle creatine kinase gene (12). If class B HLH pancreatic B cell specific protein exists, a similar model might be proposed that IEBP1 and this protein form hetero-dimer, bind to the rat insulin 1 gene enhancer and stimulate the insulin gene transcription in a tissue specific manner. It will become important to characterize these related proteins and study the interactions between these binding factors. The results presented here will help to elucidate these issues.

ACKNOWLEDGMENTS

We wish to thank Shunsuke Ishii and Toshio Maekawa for their technical advice and helpful discussion. This work was supported by a Grant-in-Aid from the Ministry of Education of Japan.

REFERENCES

1. Walker, M.D., Edlund, T., Boulet, A.M., and Rutter, W.J. (1983) Nature 306, 557-561.
2. Edlund, T., Walker, M.D., Barr, P.J., and Rutter, W.J. (1985) Science 230, 912-916.
3. Karlsson, O., Edlund, T., Moss, J.B., Rutter, W.J., and Walker, M.D. (1987) Proc. Natl. Acad. Sci. USA 84, 8819-8823.
4. Crowe, D.T., and Tsai, M.J. (1989) Mol. Cell. Bio. 9, 1784-1789.
5. Whelan, J.W., Poon, D., Weil, P.A., and Stein, R. (1989) Mol. Cell. Bio. 9, 3253-3259.
6. Lenardo, M., Pierce, J.W., and Baltimore, D. (1987) Science 236, 1573-1577.
7. Karlsson, O., Walker, M.D., Rutter, W.J., and Edlund, T. (1989) Mol. Cell. Bio. 9, 823-827.
8. Ohlsson, H., and Edlund, T. (1986) Cell 45, 35-44.
9. Ohlsson, H., Karlsson, O., and Edlund, T. (1988) Proc. Natl. Acad. Sci. USA 85, 4228-4231.
10. Moss, L.G., Moss, J.B., and Rutter, W.J. (1988) Mol. Cell. Bio. 8, 2620-2627.
11. Murre, C., McCaw, P.S., and Baltimore, D. (1989) Cell 56 777-783
12. Murre, C., McCaw, P.S., Vaessin, H., Caudy, M., Jan, L.Y., Jan, T.N., Cabrera, C.V., Buskin, J.N., Hauschka, S.D., Lassar, A.B., Weintraub, H., and Baltimore, D. (1989) Cell 58, 537-544.
13. Singh, H., LeBowitz, J.H., Baldwin, A.S., and Sharp, P.A. (1988) Cell 52, 415-423.

14. Maekawa, T., Sakura, H., Ishii, C.K., Sudo, T., Yoshimura, T., Fujisawa, J., Yoshida, M., and Ishii, S. (1989) EMBO J. 8, 2023-2028.
15. Maekawa, T., Sakura, H., Sudo, T., and Ishii, S. (1989) J. Biol. Chem. 264, 14591-14593.
16. Sakura, H., Maekawa, T., Imamoto, F., Yasuda, K., and Ishii, S. (1988) Gene 73, 499-507.
17. Landschultr, W.H., Johnson, P.F., and McKnight, S.L. (1988) Science 240, 1759-1764