Vol. 148, No. 1, 1987 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
October 14, 1987 Pages 344-349

A NEW LEADER EXON IDENTIFIED IN THE RAT INSULIN-LIKE GROWTH FACTOR II GENE

Takaki Ueno, Keikichi Takahashi, Tetsuya Matsuguchi, Hideya Endo, and Mikio Yamamoto

Department of Biochemistry, Medical Institute of Bioregulation, Kyushu University, Higashi-ku, Fukuoka 812, Japan

Received September 1, 1987

SUMMARY: In the rat insulin-like growth factor II gene, we identified an alternative 5'-untranslated sequence, in addition to the two sequences documented recently (1986) J. Mol. Biol. 192. 737-752, (1986) J. Biol. Chem. 261. 17138-17149. This sequence was encoded in an extreme 5'separate exon (E1 exon), the promoter region of which could be distinguished from others by features including lack of the TATA-box and GC-repeat, and heterogeneity in the transcriptional initiation.

The insulin-like growth factors I and II (IGFI and IGFII) or the somatomedins are mitogenic polypeptides that play an important role in fetal and postnatal growth. The presence of multiple RNA species is a common expression pattern of IGF genes (1-4). Analysis of rat IGFII (rIGFII) cDNAs and its genome led to identification of two promoters which are partially responsible for generating multiple rIGFII RNAs (1,2). We report here a detailed analysis of 5'-structures of rIGFII RNAs and the gene, and the presence of an additional leader-exon located in a further 5'-region of the recently-characterized two leader-exons (1,2).

EXPERIMENTAL PROCEDURES

Recombinant Libraries: The poly(A)$^+$ RNA was prepared from transplantable rat hepatoma AH60C cell line by a combination of guanidine thiocyanate-CsCl extraction (5) and oligo(dT)-cellulose column fractionation. A cDNA library primed with poly(dT)-tailed plasmid was prepared from AH60C poly(A)$^+$ RNA as described (6). A primer-extended cDNA library was constructed with AH60C poly(A)$^+$ RNA and synthetic oligonucleotide (primer I, 5'CCACTTCCTACCCATAGT TT 3' or primer II, 5'CGATGCAGCACAAGGCGAAG3') according to the method of Watson and Jackson (7) , except for the use of methylmercuric hydroxide for RNA denaturation.

0006-291X/87 $1.50

The rat genomic library in λ Charon 4A phage was purchased from Clontech Lab. Inc. (Palo Alto, CA.).

Other methods: Nucleotide sequence was determined by the dideoxy chain-termination method (8). Conditions for S1 nuclease analysis were essentially as described (9).

RESULTS AND DISCUSSION

We isolated several rIGFII specific cDNAs from hepatoma AH60C libraries primed with either poly (dT)-tailed plasmid or the synthetic oligonucleotide primer I. A schematic representation of rIGFII cDNA with a restriction map is shown in Fig.1A. The composite sequence consisted of 543 bases of the coding region for the rIGFII precursor protein and 2889 bases of the 3'-untranslated region. The clones from a primer I-extended library, however, contained three different sequences (E1, E2 and

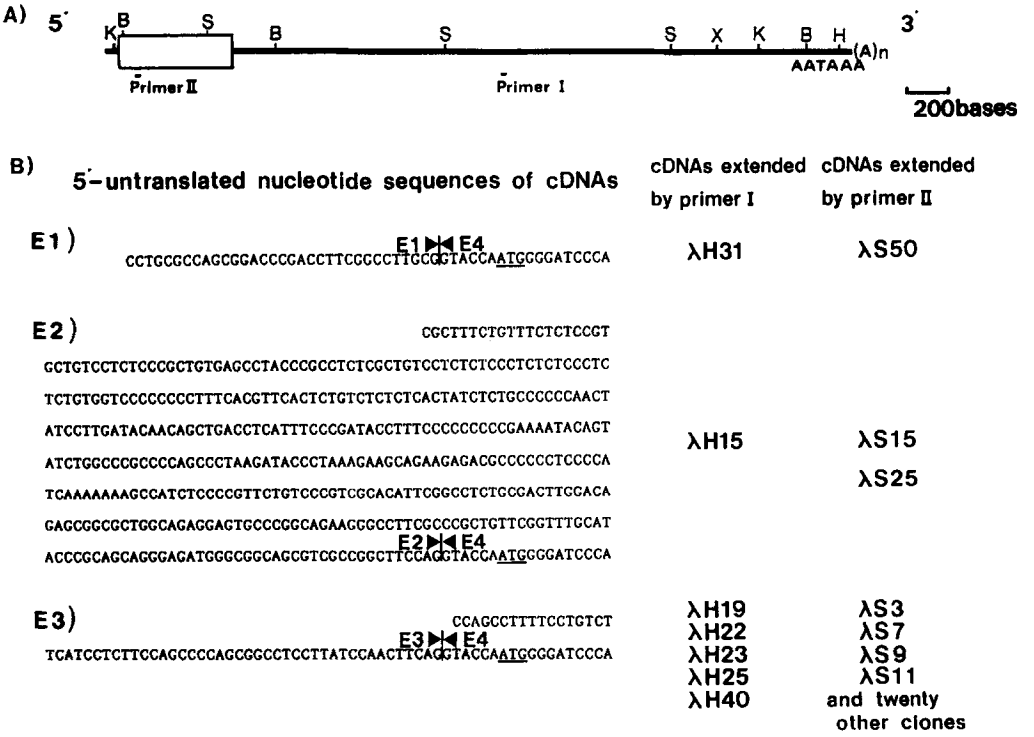


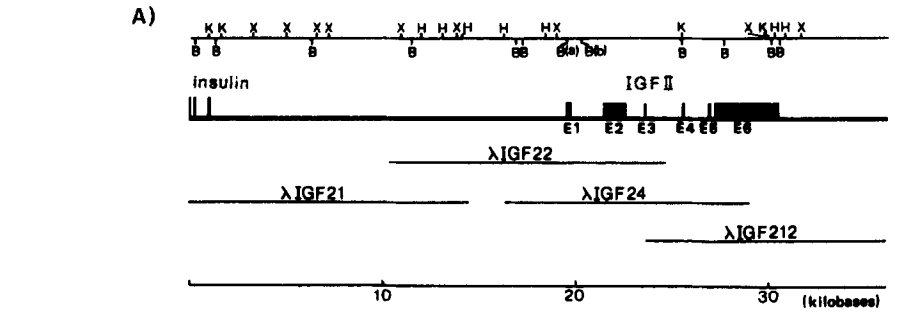
Fig.1. A) Schematic presentation of rIGFII cDNA. The coding and 3'-untranslated regions are represented by the open box and the thick line, respectively. The positions of primer I and primer II are shown under the cDNA structure. Restriction sites are : B, BamHI; H, HindIII; K, KpnI; S, SacI; X, XbaI. B) Alignment of three 5'-untranslated sequences of rIGFII cDNAs. ATG initiator was underlined. E1, E2, E3 and E4 exons are denoted in Fig.2. The vertical line indicates the exon-exon junction. The cDNA clones are classified according to their 5'-untranslated sequences.

E3) as the 5'-untranslated region diverging from the same position onwards (8 bases upstream from ATG initiator) (Fig.1B). Comparison with the genomic sequences published by Soares et al. (1) and Frunzio et al. (2) revealed that the sequences of E2 and E3 were identical to those of longer and shorter alternative 5'-untranslated regions, respectively, which were encoded in different exons. However, there was no matching sequence to that of E1.

To analyse the presence of another 5'-untranslated sequence in rIGFII transcripts, we constructed a new library by priming with the primer II locating at the 5'-proximity of the coding region (see Fig.1A). From the sequence analysis of 30 independent rIGFII clones isolated from this library, one clone had a homologous sequence to E1, while twenty-four and two clones were identical to E3 and E2, respectively (Fig.1B). The remaining three contained only the coding region. No additional 5'-untranslated sequence was observed other than three variants. These results indicate that three alternative 5'-untranslated regions are present in rIGFII mRNA, and the RNA species with E1 and E2 sequences are about 10-fold less abundant than that with the E3 sequence.

Four overlapping recombinant phage clones were isolated from a rat genomic DNA library to determine the location of the E1 sequence on the rIGFII gene. The complete map of the rIGFII genomic region is shown in Fig.2A. The contiguous organization of the rIGFII and the rat insulin II genes (10) was confirmed by hybridization of λIGF-21 clone with the rat insulin cDNA. The sequence homologous to E1 was found 1.4 kilobases(kb) upstream from E2 exon, and the 3'-end of the sequence contained a canonical donor splice site (Fig.2B). Probes prepared from the upstream region of the BamHI(a) site gave no positive signal on Northern analysis (data not shown). The E1 exon was thus considered to be located at the 5'-extreme of the rIGFII gene, which consisted of six exons (E1 through E6 exons from the 5' to 3'). Exons E1, E2 and E3 encoding 5'-untranslated region were used alternatively, and E4 and E5 exons contained the major part of the coding region. The 3'-part of the coding region and the whole 3'-untranslated region were located in a single exon E6.

To delimit the E1 exon, S1 nuclease analysis was performed. The positions of the 5'-ends were determined with an AvaI-AvaI probe (45 to 757) as shown in Fig.3. Four major doublet and two



B)

XbaI Ava I

TCTAGAGGACTTCTCTGAAGCCACAGAAATTAGAGGTGACTTTTCCCGGGTGCTGTAGGGCCCACTGGCTTCAGGGTGAAACTTAGCCTCCACTCCCACC 100

CCCAAAAGAATTGACCCTTAATTCCCTAAAACCCATCTGTTGTTTTTTTCACTCTCCCATTTTTAGGCAATTAAAATTAGCTCCCCCCCCCCAAATTTCA 200

TGTTGGGAATTTTTCTTAAACTCAACAGCAACAGCAAACCACCTAACTTCTATAAAATAGCCTAAGATTTTGTGGGTTCCCCAAACTTGCTGACTAAGACC 300

TTAGGTTCCCCACGTTAGGCTTGGATCAAGATGTTCCCCTTTTTTGCATAGAATAATTTTCCTCATTTCCCAACTTGGATTCTAATTTTTCTGGAATCTC 400

AAATTTAGTCGTTTGCCACAGTTTTCCCATTTATGGAGTCTCATTCAAGTCCTCCTACTTTGGGCCACTCATTCCCAGGGCCGCAAATCAGACAAGGGTC 500

BamH I(a)

TGATTTGGGATCCCCAAATTTGATCATTGTTTCCGCTCCCCACTTTCGCCAGCTTCCCGTCCTCCCCATAACATCCATATCCCCCTCTTCACTTCTCCT 600

ACGGTGTCCCAGCCGCCAGCTCAGATTTTGCAAAAGTGTGTGTCCCCAAACACGCACCGGACACTGCGCGGCGCGCCAGGACGTGGGCAGTGCTGCTCCCG 700

Ava I

TGTCCAGGAAAACCACTGGGCATTGCCCCCAGTTTCCCCCAAATTTGGGCATTGTCCCCGGGTCTTCCAACGGACTGGGCGTTGCCCCGGGACACTGGGG 800

Ava II

ACTGCCTTTGGGGTCTCGCTCACCTTCAGCAACGTCCACTTCAGCAGCTCCCACTTCTGCAGCTCTCCCGCCCTTCTCCGATCCTCCTGCGCCAGCGGAC 900

CCGACCTTCGGCCTTGCGGTGAGCTCCCGGTGGCGCGCTCCCTTGTGCTCCAGCTCTTCCGTTGCGGGGCACCGGCTACCTCGCAGTTTGTCTTCAGGGC 1000

TGCTGCGGGGTAGCGCGGGAGCGGGCTGCTGGCTTCTCCCTTGGCGACCGCGCTCCTTTGGCCGGCGCTCCGAAACACGCGCCACTTTGTATGCACTC 1100

BamH I(b)

TGCAAACAACGCGGTGGTAGGGAAAGCGCCTGAGCGGCCAGGAAGCGGGAAAGGCAGAGCGCTCCGGCGGCTTGGATCC

Fig.2. rIGFII gene structure and nucleotide sequence around exon E1. A) Combined restriction map and exon-intron organization of rIGFII gene. Restriction sites are as in Fig.1. Exons are represented by filled boxes, and introns and untranscribed regions are shown by the thick line. The genomic DNAs are derived from four overlapping recombinant phages shown underneath. B) Nucleotide sequence of the E1 exon and flanking regions. The sequence corresponding to E1 cDNA sequence in Fig.1B is double-underlined. 5'-boundaries of E1 exon are marked with asterisks. A pair of horizontal arrow represents a dyad symmetrical sequence, and ATTTGGG and AATTTTCC repeats including their variants are boxed and underlined, respectively.

minor triplet bands (marked with asterisks in Fig.2B) were observed, and the intensities increased proportionately with the amount of RNA and persisted against increased concentrations of S1 nuclease. This 5'-end heterogeneity was also observed by S1 analysis using the BamHI(a)-AvaII fragment (508 to 897) (data not shown). The length of E1 exon varied thus from 260 to 371 bases. The primer extension analysis using a 39 mer complementary to nucleotides 686 to 724 yielded products extended up to the

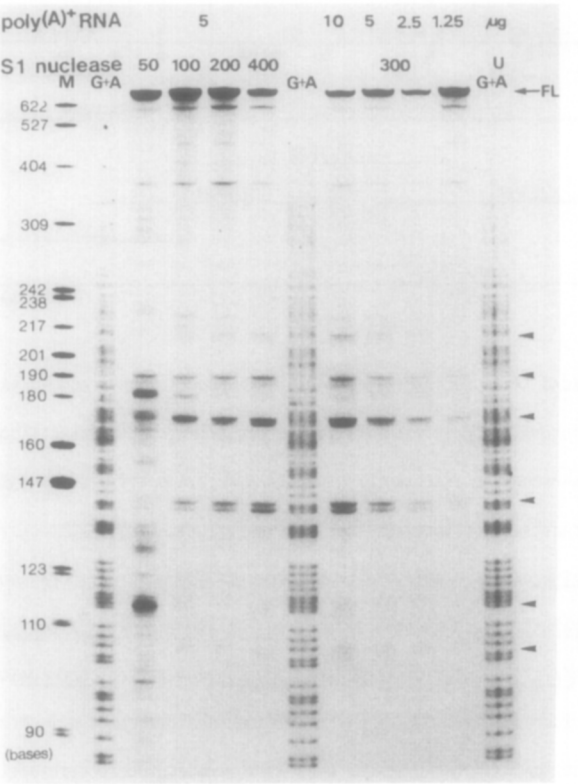


Fig.3. S1 nuclease analysis of the E1 exon. A 712 bases AvaI-AvaI fragment was labeled at 5'-end. After separation in a single strand, the probe was hybridized with indicated amounts of poly(A)$^+$RNA (1.25 to 10μg), then subjected to S1 nuclease (50 to 400 units) digestion. Arrowheads indicate protected bands. FL: full length probe. G+A: G+A sequence ladder. M: MspI-fragments of PBR322.

similar positions observed with S1 analysis (data not shown). These results indicate that the initiation of E1 transcription takes place heterogenously at multiple positions. Northern analysis using E1 probe and AH60C poly(A)$^+$RNA gave major band at about 3.8kb in length, while E2 and E3 probes gave 4.6kb and 3.6kb major products, respectively (data not shown). These sizes were in good agreement with those expected from cDNA and genomic sequences.

The promoter region upstream of E1 exon (E1 promoter) was distinguished from those of E2 and E3 exons by several remarkable features. Firstly, typical eukaryotic promoter elements including TATA-box and GC-repeat were present in the E2 and E3 promoters (1,2) but absent in the E1 promoter. The lack of an apparent TATA-box sequence must be related to the heterogenous startsites, as observed in other TATA-less genes (11,12).

Therefore, it is of interest to note that a number of dyad symmetrical sequences are located at positions 20 to 30 bases preceding many startsites. These sequences may mimic the TATA function. In addition, the intergenic region contains two repeats, ATTTGGG and AATTTTCC. The former repeat is homologous to a CAAT box in dihydrofolate reductase gene (13). Secondly, a number of short AT rich tracts were scattered in regions far upstream from the startsites. Some of these sequences were closely related to the cleavage consensus of topoisomerase II which is sometimes associated with enhancers (14).

ACKNOWLEDGEMENTS

We thank Dr.G.I.Bell for the generous gift of rat insulin cDNA, Dr. Y. Sakaki and M.Ohara for comments on the manuscripts and N.Kinoshita and N.Mochizuki for technical assistance. This work was supported in part by Grants-in-Aid from the ministry of Education, Science and Culture, Japan, and by the research grant from the Japan Tobacco Industrial Co.

REFERENCES

1. Soares,M.B., Turken,A., Ishii,D., Mills,L., Episkopou,V. Cotter,S., Zeitlin,S. and Efstratiadis,A. (1986) J. Mol. Biol. 192, 737-752.
2. Frunzio,R., Chiariotti,L., Brown,A.L., Graham,D.E., Rechler,M.M. and Bruni,C.B. (1986) J. Biol. Chem. 261, 17138-17149.
3. Rotwein,P. (1986) Proc. Natl. Acad. Sci. USA 83, 77-81
4. Scott,J., Cowell,J., Robertson,M.E., Priestley,L.M., Wadey,R., Hopkins,B., Pritchard,J., Bell,G.I., Rall,L.B., Graham,C.F., and Knott,T.J. (1985) Nature 317, 260-262
5. Chirgwin,J.M., Przybyla,A.E., McDonald,R.J. and Rutter,W.J. (1979) Biochemistry 18, 5294-5299.
6. Okayama,H and Berg,P. (1982) Mol. Cell. Biol. 2, 161-170.
7. Watson,C.T. and Jackson,J.F. (1985) in : DNA cloning (Glover D.M.ed) vol.I pp79-88, IRL Press, Oxford, Washington DC.
8. Sanger,F., Nicklen,S. and Coulson,A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467
9. Barrera-Saldana,H., Takahashi,K., Vigneron,M., Wildeman,A., Davidson,I. and Chambon,P. (1985) EMBO J. 4, 3839-3849
10. Cordell,B., Bell,G., Tischer,E., DeNoto,F.M., Ullrich,A., Pictet,R., Rutter,W.J. and Goodman,H.M. (1979) Cell 18, 533-543
11. Fiers,W., Contreras,R., Haegeman,G., Rogiers,R., Van de Voorde,A., Van Heuverswyn,H., Van Herreweghe,J., Volckaert,G. and Ysebaert,M. (1978) Nature 273, 113-120
12. Cowie,A., Tyndall,C. and Kamen,R. (1981) Nucl. Acids Res. 9, 6305-6322
13. McGrogan,M., Simonsen,C.C., Smouse,D.T., Farnham,P.J. and Schimke,R.T. (1985) J. Biol. Chem. 260, 2307-2314
14. Gasser,S.M. and Laemmli U.K. (1986) Cell 46, 521-530