Identification of a novel transcription unit in the human insulin-like growth factor-II gene

Koji IKEJIRI,* Taro WASADA,† Kosuke HARUKI,† Naomi HIZUKA,‡ Yukimasa HIRATA† and Mikio YAMAMOTO*§

*Department of Biochemistry, Medical Institute of Bioregulation, Kyushu University, 3-1-1 Maidashi, Higashi-ku, Fukuoka 812, Japan, †Diabetes Center and ‡Institute of Clinical Endocrinology, Tokyo Women's Medical College, 8-1 Kawada-cho, Shinjuku-ku, Tokyo 162, Japan, and §Second Department of Biochemistry, National Defense Medical College, 3-2 Namiki, Tokorozawa, Saitama 359, Japan

The human insulin-like growth factor-II (hIGF-II) gene has until now been thought to be composed of eight exons, including three independent leader exons. In the present study two additional exons, one leader exon and one alternatively used ordinate exon, have been newly identified. They were abundantly expressed in human histiocyte tissue, generating mRNA species of about 5.0 kb in length. The new leader exon shows significant sequence similarity with the rE1 exon, previously reported to be transcribed only in the rat, and is mapped at nearly the same genomic location as in the rat. On the other hand, sequence similarity with another exon in the corresponding region of the rat genome was also found. It was, however, obvious that the rat sequence would not work as an active exon, since both splice acceptor and donor sites were deviated considerably from the consensus sequences. It has thus become apparent that the complex transcription unit of a single-copy hIGF-II gene comprises at least 10 exons, including four leader exons, one alternative exon and three common protein-coding exons.

INTRODUCTION

Insulin-like growth factor-II (IGF-II) is a polypeptide hormone related to insulin and IGF-I in both structure and function [1,2]. The IGFs evoke a number of biological activities, including cellular induction of DNA synthesis, protein synthesis and cell multiplication [3–6]. IGF-I is thought to mediate the growth-promoting effects of growth hormone [7]. Although the function of IGF-II is less well defined, it has been proposed that it serves as an important factor in fetal differentiation and development, since levels are high in fetal plasma but decline rapidly in the postnatal period [8], and possibly in the function of the central nervous system [9–13]. High levels of IGF-II mRNA are found in most human and rodent fetal tissues, whereas in the adult the same tissues show much less, if any, expression [9,11–16]. IGFs are also secreted by various tumours and cultured cell lines [10,14,17–22]. They may thus play an important role in the deregulation of growth control in certain types of tumours, perhaps through autocrine or paracrine mechanisms [23,24].

A common transcriptional feature of the IGF family is the generation of multiple RNA species, despite the copy number of one for the encoding gene. Major species of these RNAs extend up to 4–6 kb, whereas only about 400–600 nucleotides are sufficient to code for the entire peptide [9,10,21,25,26]. Structural analysis of rat and human IGF-II cDNA clones and their genes reveal the presence of alternatively used leader exons which are responsible for the generation of multiple transcripts, in that they provide 5'-untranslated regions that are variable in length and sequence. Up until now three independent transcriptional promoters have been identified in each species, two of which are closely related in genomic location and sequence. The third (5'-extreme) one is, however, specific to each species and differs not only in these criteria but also in their transcriptional regulation [27–29].

We have recently determined the entire genomic sequence of rat IGF-II (rIGF-II), including the 5' upstream intergenic region up to the neighbouring insulin gene and the 3' region of 6 kb from the major polyadenylation site [30]. By comparing this sequence with reported human exonic sequences, we found the complete absence of a human-specific leader exon h1 analogue in the rat genome. To determine whether a sequence similar to the (so far) rat-specific exon rE1 is also present in the human genome and is expressed, we screened mRNA preparations from several human tissues by Northern blotting using a probe specific to the rE1 sequence. We found abundant expression of an rE1-like sequence in human histiocyte mRNA. In the present study we analysed cDNAs obtained from this tumour line containing the rE1-like sequence. Human genomic regions related to this mRNA class were also determined and compared with the corresponding genomic portions from the rat.

MATERIALS AND METHODS

Northern blot analysis

Poly(A)⁺ RNAs were isolated from adult human liver and histiocyte as previously described [31]. mRNA (3 µg) was electrophoresed through a 1.5% (w/v) agarose/6% (v/v) formaldehyde gel and blotted on to nitrocellulose paper, according to the method of Thomas [32]. The paper was then baked at 80 °C for 2 h, prehybridized and hybridized with the DNA fragment which had been ³²P-labelled by random priming [33].

DNA sequence analysis

Nucleotide sequences were determined by the dideoxy chain-termination method of Sanger *et al.* [34], with slight modifications [35], after subcloning of DNA fragments into pUC vectors [36]. Sequences were determined on both strands and all of the

Abbreviations used: IGF, insulin-like growth factor; the prefixes h and r denote human and rat respectively.

§ To whom correspondence should be addressed.

The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under accession numbers X56539 and X56540.

subcloned fragments were overlapped by the use of more than one restriction enzyme.

Construction of a cDNA library

The cDNA library was constructed against poly(A)⁺ RNA from histiocytoma tissue curatively excised from the retroperitoneum, according to the method described by Gubler & Hoffman [37]. *Eco*RI linker-ligated double-stranded cDNAs were inserted into the λ zapII vector (Stratagene, La Jolla, CA, U.S.A.). The final population size was 5×10^6 clones, and was amplified once before screening.

Genome analysis of hIGF-II

A genomic library was constructed using *Mbo*I-digested DNA from the term placenta from a Japanese subject, *Bam*HI-digested EMBL3 DNA (Stratagene) as a vector and a Gigapack Gold *in vitro* packaging kit (Stratagene). Two independent clones containing the hIGF-II genome were obtained by plaque hybridization using a hIGF-II-coding region-specific probe. A series of *Sal*I subclones was generated for subsequent DNA sequence analyses.

Primer extension

Primer extension was performed essentially by the established method [38], for the preparation of labelled primer. About 0.5 pmol of a 23-mer oligodeoxynucleotide was labelled at its 3' downstream portion after annealing with 1 pmol of single-stranded phagemid containing the genomic hIGF-II-h4 region and with a T7 DNA polymerase sequencing kit (Pharmacia LKB Biotechnology, Uppsala, Sweden) according to the procedure recommended by the manufacturer. After completion of the labelling reaction with [α -³²P]dCTP and successive heat denaturation, high-molecular-mass template and unused dCTP were removed by differential propan-2-ol precipitation and ethanol precipitation respectively. Poly(A)⁺ RNA (10 μ g) was annealed with about 4×10^4 c.p.m. of labelled primer and then reverse-transcribed.

RESULTS

The rat leader exon rE1 sequence has been reported to be expressed only in the rat [39]. In order to determine whether a similar sequence is also expressed in the human, we examined an mRNA sample from normal adult human liver by Northern blotting, using as a probe the rat genomic region representing rE1. As shown in Fig. 1(a) (lane 1), such a sequence was not detected. However, the sequence analogue was found to be expressed abundantly in a histiocytoma which occurred in the retroperitoneum (Fig. 1a, lane 2). The major species of mRNA was estimated to be about 5 kb in length. The Northern blotting pattern was nearly identical to that obtained with the rIGF-II-coding region-specific probe (Fig. 1b, lane 2), indicating that rE1-positive signals were indeed derived from hIGF-II mRNAs. Most IGF-II mRNA species from human histiocytoma thus appeared to contain an rE1-like sequence. The rIGF-II-encoding region-specific probe also generated a single band in the sample from normal adult liver (Fig. 1b, lane 1), which is considered to be due to the transcript from the adult specific leader exon h1 [28]. The same RNA blot was re-hybridized with human β -actin probe to examine the integrity and amounts of RNA samples from both tissues (Fig. 2c).

To elucidate the primary structure and exon arrangement of the hIGF-II mRNAs containing the rE1-like sequence, we prepared a cDNA library from the tumour mRNA and screened it with the rE1 probe. About 0.05% of the clones revealed a

positive signal, indicating moderate abundance of the sequence in this particular tumour tissue. Thirty clones were randomly chosen, and seven clones containing longer inserts were selected for complete sequence determination. Unexpectedly, they could be divided into three types with respect to their possession of specific sequence stretches. Fig. 2 shows these three types of cDNA containing the rE1-like sequence by alignment with the genomic sequences determined at the same time. The region highly analogous to the rE1 sequence was in the 5' portion (upstream of nucleotide 182; see also Fig. 4). Type a mRNA contained the sequence stretch of nucleotides (−181 to −7) and a shorter stretch of nucleotides (+178 to +186), whereas types b and c lacked the latter part and both parts respectively. Sequence regions of nucleotides −6 to +177 and +187 to +358, and the region downstream of nucleotide +359, correspond to hIGF-II exons 5, 6 and 7 respectively [27]. These results clearly show the presence of a new transcription unit in the hIGF-II genome, the second exon of which is occasionally skipped by an apparent alternative splicing mechanism. Occasional inclusion of the short sequence stretch (nucleotides 178 to +186) into the mRNA has been reported before [26], from analysis of cDNAs prepared from adult liver mRNA. Since this alternative splicing acceptor selection occurs in the middle of the coding region, the inclusion of this region would result in the replacement of the Ser residue with Arg-Leu-Pro-Gly in the B domain of the mature peptide. Of particular note was the genome sequence around the alternative splicing site (nucleotides +176 and +177), where a GG sequence occurred instead of AG. To examine the possibility that the sequence of this particular position might be AG as an allelic variation, we extracted DNA from the same histiocytoma tissue,

Fig. 1. Northern blot analysis of human histiocytoma transcripts containing the sequence hybridizable to the rat rE1 exon sequence

Poly(A)⁺ RNAs (3 μ g/lane) from normal adult liver (lanes 1) and histiocytoma (lanes 2) were processed as described in the Materials and methods section. RNA blots were hybridized with probes representing rat rE1 (a), the rat IGF-II coding region (b) and human β -actin (c). Triangles denote the positions of 28S and 18S rRNAs. Numerals at the left indicate the sizes, in kb, of the major hIGF-II mRNAs reported previously [27].

Human histiocytoma cDNA clones containing the sequence cross-hybridizable to the rat rE1 sequence were isolated, and seven longer sequences among 30 randomly chosen clones were determined. Genomic clones were also isolated using the same criteria, and portions corresponding to cDNA sequences were determined. Values in parentheses at the 3' ends show the numbers of cDNA clones sequenced. Nucleotides are numbered sequentially from the transcriptional initiation codon ATG (boxed) of the human IGF-II precursor. Triangles denote exon-intron junctions.

last part of exon h4 is only slightly similar to that of rE1, the exon-intron junction corresponded perfectly with that for the rat [39], and the sequences around the splice donor site are also well-conserved in both species. Especially high similarity (95%) was apparent in the central one-third portion (nucleotides +536 to +637), whereas both the upstream and downstream regions of this part were less similar (Fig. 4). It is thought that human exon h4 represents a leader exon for the following reasons: (1) its sequence resembles that of the rat leader exon rE1, and (2) no distinct signal was detected on Northern blots using the genomic sequence upstream from this sequence to exon h3 (results not shown). Since transcriptional initiation in the rat rE1 exon has been proved to be highly heterogeneous [39], similar heterogeneity might be expected in human exon h4. In the human exon, however, it is obvious that the transcription could start from a region further upstream than the extreme 5' site established in the rat genome (marked with the symbol $\bigcirc$ in Fig. 4) [39], because of the presence of type a mRNA (5'-most asterisk in Fig. 4). The transcriptional start site was therefore surveyed by the primer-extension technique (Fig. 5). Multiple extended bands were obtained around nucleotides +320 to +420, as predicted. Some putative start sites were indeed located further upstream from the site where the type a cDNA started.

Based on the general similarity of the genomic arrangement in the rat and human (Fig. 3), we searched the rat genome for a sequence stretch similar to the newly identified human exon h5. Since we have determined the entire nucleotide sequence of the rat IGF-II gene region, including the entire intergenic region up

Fig. 3. Schematic representation of rat and human IGF-II gene regions and their transcriptional patterns

Boxes indicate exons. Human exons are numbered sequentially from 5' to 3'. The pattern for the rat was taken from a previous report [39]. * indicates the presence of a variant mRNA due to alternative splicing. Ins represents the insulin gene.

Fig. 5. Primer extension analysis

A 23-mer oligodeoxynucleotide complementary to the indicated part of the mRNA (AAGCCCTCCCTGTCCACGTCCTG) in Fig. 4 was labelled as described in the Materials and methods section, hybridized to histiocyte (H) or liver (L) mRNAs. For lanes A, C, G and T, the sequencing reaction was performed on genomic h4 region with the same 23-mer oligodeoxynucleotide. Although the results were obtained for the complementary strand, the Figure was put face down in order to see the mRNA strand. Since the same primer was used for primer extension and sequence analysis, it was possible to assign exactly extended bands to the genomic region. Putative transcriptional start sites were marked on Fig. 4 (●).

h4	GGATCCCCAAAATGTGTTCTTCTTTCAT-CTGCCAATTTACGTAA-TATGGCTCTACGGCAAAATGCCAATTCATATGGAGAATTTCTTTAACT	98
rE1	TAAAATTAGCTCCCCCCCCCAAAATTTCTGTTGGGAATTTTCTTAACTCAACAGCAACAGCAACCACTTAAGTTC-TATAAAATAGCCTAAGATTTT	99
h4	ACCCCTCCTCACAATTTGGTCCCAAGCTAGCTGAGCCCTATTTGAGACCTCTTTCTCTATGTTCCG---AATTGCATGGAGCAACTTCTGTCATCCGC	195
rE1	GTGGGTTCCCAAACTTGTGCTGACTAAGACCTTAGGTTCCCAAGTATAGGCTTGGATCAAGATGTTCCCTTTTTCATAGAATATTTCTCTATTTCG	199
h4	CAAACTGTAAATCT-ATTTTCTGGAGTCTCAGTATTAGTCATTAATCAGGTTCCCAATTAACGGAG--TCCCGGGGTCC-CCTCTCCAGGACACC	291
rE1	C-AACTTGGATTCTAATTTTCTGGAATCTCAAATTTAGTCGTTTGGCAGATTTTCCCATTTATGGAGTCTCATTCAAGTCTCTCTACTTGGGGCACT	298
h4	CATTGCTAAGCCCGCAAGGCAAGAACTCTGCCTTTCGCTTCCCAAAATTTGGGCATTGTTCCGGCTCGCCGGCCACCCACTGAGCTTGGCAACC	391
rE1	CATTG-CCAGGGCCGCAAACTCAGACAAGGCTCTGATTGGGATCCCGAAATTTGATCATTGTTTCCGCT---GGCCACTTTCGCAAGCTT-CGGGT-C	391
h4	CGCGCACAGCGGCACTGGTTTCGGGC-C-TCTCTGTCTCCTACGAAGTCCCGAGAGCACTGGGATTGGGAAATTTCTCTCTAGCGTTGCCAAACA	489
rE1	CTGCGCAT----ACATCGATATCGGCCTCTTCACT-TCTCTACGGGTGTCAGAGCGGCACTCAGATTGGGAAAGTGTGTGT---GCCCAACA	481
h4	CACTTGGGTCCGGCCGCGC-GCCCTCAGGACGTGGACAGGAGGGCTTCCCGGTGTCCAGGAAAGCGACCGGGCATTGCCCCCACTCTCCCAAAATTTGG	588
rE1	CGCACTCGGAC-ACTGCGCGGCGGCGGACGCTGGGACGTG-CTGCT---CCCGTGTCCAGGAAAGCACTGGGCAATGCCCCCACTTCCCAAAATTTGG	577
h4	GCATTGTCCCGGGTCTTCCAACGACTGGGCGTTGCTCCCGGACACTGAGGACTGGGCGGGGCTCGCTCAGCTTCAGCAGCGTCCAC--CGCCTGC	686
rE1	GCATTGTCCCGGGTCTTCCAACGACTGGGCGTTGCTCCCGGACACTGAGGACTGGGCGGGTCTCGCTCAGCTTCAGCAGCGTCCAC--CGCCTGC	677
h4	-CACAGAGCGTTTCGATCGCTCG-CTGCTGAGCTCTGGTGGCGCGGCGGACGAGCTCCAGCTTCGCTGAGCTCCCGCGCGCGATCCCTCCG	781
rE1	TCCCACTTCTGACGCTCTCCCGCCTTCTCCGATCTCTCTGCG-CCACGGAGCCGACCTTGGGCTTGGCTGAGCTCCCGGCGGCG-CTCC-CTCC	770
h4	CCTCTGCGCCCTGACCGGCTCTCGGCGCGCATCTGCTGCTGCTCCGCGCGGTGCTGGCGCTCGTCTCCGCTGCGCGCGGGGAGGCGGCTGGGCGCGG	881
rE1	-CT-TGTGCTCCAGCTCTTCCGTTGCGGGGACCGGCTAC---CTCGCAGTTGT-----CTTC-AGGCTGCTGC---GGG-----GTAGCGGG	817
h4	GACACGGCTGCGGACTT-GCGGCTGGCGGCTGCGCTGCTGCTGGGCGCGCGAAATCCGCGGACCACTTTGTTGCTCATTTGCAAAAGATCTCATT	983
rE1	GAGCGGGCTGCTGGCTTCTCCCTTGGCGACCGGCTCTTGGGCGGGCGGCTCCGAACACGCG--CCACTT-GTATGCACTCTGCAAAACACGGGT	944

Fig. 4. Comparison of the rat genomic sequence corresponding to the rE1 region and that of the newly identified human h4 region

Transcriptional initiation sites identified in the human (see Fig. 5) and the rat [39] are indicated by ● and ○ respectively. Colons (:) between two sequences indicate the same nucleotide. Asterisks denote the 5'-termini of cDNA clones analysed. A cluster of small arrows indicate the stretch where the anti-sense oligodeoxynucleotide primer was prepared. Gaps have been introduced to maximize alignment.

h5	TGACGTGGT TTTT GACTTGGTAAAAATCACAGTCTTTCTTACATCGTTCAAACCTCCAGAGATGGTTTCCCCAGACCCCAAAATTATCGTGGTGGCC	100
r	AAGAGTGA-TTTT GACTTGGTAAAAACTCAATG-GC-TTCTTATATC--TGAAACCCACC--GAGATAGCTTCTTAACAACCCCAAAATTATCCT-GTTGCC	92
h5	CCCGAGACCGAAGCTCGGCTATGCAAGTCCAACGCACTGAGGACGGGGTAACCATTTCCAGATATTTTGGGTGGGCGCAAAGGCGAGCTACTTAGAC	200
r	CCTGGGGCGGAACCTGCAACTAAGCAAATTCAAAGTGCTAAGGACAGGG-AACTGTTATCTGGACCTTCTGGGTGGGTCCAAAGGCGCGCGCTGAGAA	191
h5	GCACCCCGGTGAGCTCGGCCATGCACTAGGATTTGAGCTGTGTTACCGCCCTGATCCC	260
r	GCAATTTAGTGAGCTCTGGCCCCAGGCTCTGCTTTGAGCTACGATTCCCGCCCTGCTTC	251

Fig. 6. Comparison of the human genomic region representing exon h5 and a similar sequence from the rat

The position of exon h5 is indicated by arrows. Those of rat (r) sequences corresponding to the human splice acceptor and donor are underlined. The rat sequence was taken from our previous report [30].

to the 5' neighbouring insulin gene [30], it was possible to search for a similar region in the rat genome and to make an alignment with the human exon h5 region. As shown in Fig. 6, a similar sequence was indeed found in the rat genome at a location corresponding to that in the human genome. The general sequence similarity in the exonic region was about 70%. The splice donor sequence in human exon h5 agreed well with that of consensus sequences [43]. The presence of trimeric A residues in the acceptor sequence, however, may be responsible for an occasional skip of this exon, giving rise to the generation of type c mRNA (Fig. 2). On the other hand, AG dinucleotides were deleted from the acceptor site of the rat gene, indicating the unlikelihood of the usage of this region as an active exon. Furthermore, the splice donor site was also considered to be less likely to be used, as the sequence (GTCTG) deviated markedly from that of the consensus (GTRRR) [43]. Thus it is clear that the rat sequence similar to human exon h5 could not work as an independent exon. This was also supported by Northern blotting analysis using the probe representing the region similar to h5, where no positive signal was observed in any rat fetal or tumour tissues [29].

DISCUSSION

The IGF-II gene produces multiple transcripts of various sizes partly because of the usage of multiple independent leader exons, although the gene is present as a single copy in the chromosome [27]. In the present study we have shown the presence of an additional leader exon, h4, and its ordinate but independent exon, h5, in the human IGF-II genome. Although exon h5 functionally resembles exons h2 and h3, in that they are ordinate exons and constituents of the 5' non-coding region of the mRNAs, it has a feature distinguishing it from any other exons reported so far, in the IGF-II gene, in that it is sometimes spliced out (Figs. 2 and 3). A conspicuous feature of this h4-5-8-9-10 transcription unit (Fig. 3) is its transcriptional activity, which was the highest among all four units identified in the hIGF-II gene, at least in a certain type of tumour cell. This contrasted greatly with the relative usage observed for the three leader exons in the rat, where the promoter activity was usually the highest in rE3 and the lowest in rE1 in all fetal tissues and most tumour cells [44]. Similar results were also obtained *in vitro* using transient expression assays of CAT constructs containing truncated promoter regions [45,46]. High promoter activity of the human exon h4 was therefore unexpected, since, as shown in Fig. 4, the putative promoter regions of h4 and rE1 contain similar structural features, including similar genomic map positions, partial sequence similarity and the absence of TATA and CAT boxes. Northern blotting using two other leader exon probes representing h6 and h7 (previously exons h4 and h4B respectively) also gave single bands in mRNA preparations from histiocytoma, although the signals were rather weak (M. Yamamoto & T. Wasada, unpublished work). These two leader exons are

thought to be developmentally regulated and fetus-specific [28]. Leader exon h4, co-expressed in histiocytoma together with exons h6 and h7, was thus thought to be a fetal type, although its expression in fetal tissues has yet to be identified. The transcript was not observed in adult human liver (Fig. 1a, lane 1). From the following two observations in histiocytoma, it was obvious that exon h4 transcription formed the majority among all hIGF-II transcription units: (1) Northern blotting patterns obtained with probes specific to h4 and the coding region were nearly identical (Fig. 1), and (2) more than 90% of cDNA clones having 5' non-coding regions were shown to contain the h4-h5 sequence (M. Yamamoto & K. Ikejiri, unpublished work).

Elevated expression of the IGF-II gene is a rather common phenomenon observed in many types of human tumours [10, 14, 17-22]. A contribution to deregulated overgrowth of tumour cells via an autocrine mechanism has been assumed. The rIGF-II gene has indeed been shown to be re-activated at an early stage of the 3'-methyl-4-dimethylaminoazobenzene carcinogenesis process and to have a tendency to increase its level of expression as the period of carcinogen treatment increases [44]. In this case the pattern of relative leader exon usage resembled that observed in normal embryonic tissues. In the rat embryo, IGF-II peptides are synthesized in many organs and tissues and are believed to act as embryonal growth and/or developmental factors (see the Introduction section for references). IGF-II mRNA has been shown to be localized exclusively in connective tissues or cells of mesenchymal origin by *in situ* hybridization histochemistry [47]. Since these types of cells are widely distributed and anatomically integrated into most tissues, growth of the whole embryo would be regulated by IGF-II products through the paracrine mechanism. Recently, DeChiara *et al.* [48] have shown directly the contribution of the IGF-II gene in fetal development using a combination of gene targeting and transgene techniques [48]. A similar suggestion was also made for human breast tissue and breast tumours [49] using RNAase protection analysis of IGF-II mRNA. Such analysis using leader-exon-specific probes has not yet been performed. Exon h4 expression may be low in most embryonal and tumour tissues other than histiocytoma. Although the reason for the striking enhancement of h4 expression is still unclear at present, both the *cis*-element relating to the h4 promoter and the *trans*-activating factors in this tumour tissue might be relevant. Promoters of the three leader exons in the rat IGF-II gene have different primary structures, and can be regulated independently [29].

We thank Dr. M. Furuchi for useful advice.

REFERENCES

1. Rechler, M. M. & Nissley, S. P. (1985) *Annu. Rev. Physiol.* **47**, 425-442
2. Froesch, E. R., Schmid, C., Schwander, J. & Zapf, J. (1985) *Annu. Rev. Physiol.* **47**, 443-467

3. Rechler, M. M., Podskalny, J. M., Goldfine, I. D. & Wells, C. A. (1974) *J. Clin. Endocrinol. Metab.* **39**, 512–521
4. Rinderknecht, E. & Humbel, R. E. (1976) *Proc. Natl. Acad. Sci. U.S.A.* **73**, 2365–2369
5. Froesch, E. R., Zapf, J., Audhya, T. K., Ben-Porath, E., Segen, B. J. & Gibson, K. D. (1976) *Proc. Natl. Acad. Sci. U.S.A.* **73**, 2904–2908
6. Zapf, J., Schoenle, E. & Froesch, E. R. (1978) *Eur. J. Biochem.* **87**, 285–296
7. Van Wyk, J. J., Underwood, L. E., Hintz, R. L., Clemmons, D. R., Viona, S. J. & Weaver, R. P. (1974) *Recent Prog. Horm. Res.* **30**, 259–318
8. Moses, A. C., Nissley, S. P., Short, P. A., Rechler, M. M., White, R. M., Knight, A. B. & Higa, O. Z. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 3649–3653
9. Frunzio, R., Chiariotti, L., Brown, A. L., Graham, D. E., Rechler, M. M. & Bruni, C. B. (1986) *J. Biol. Chem.* **261**, 17138–17149
10. Soares, M. B., Turken, A., Ishii, D., Mills, L., Episkopou, V., Cotter, S., Zeitlin, S. & Efstratiadis, A. (1986) *J. Mol. Biol.* **192**, 737–752
11. Lund, P. K., Moats-Staats, B. M., Hynes, M. A., Simmons, J. G., Jansen, M., D'Ercole, A. J. & Van Wyk, J. J. (1986) *J. Biol. Chem.* **261**, 14539–14544
12. Graham, D. E., Rechler, M. M., Brown, A. L., Frunzio, R., Romanus, J. A., Bruni, C. B., Whitfield, H. J., Nissley, S. P., Seelig, S. & Berry, S. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 4519–4523
13. Brown, A. L., Graham, D. E., Nissley, S. P., Hill, D. J., Strain, A. J. & Rechler, M. M. (1986) *J. Biol. Chem.* **261**, 13144–13150
14. 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. & Knott, T. J. (1985) *Nature (London)* **317**, 260–262
15. Soares, M. B., Ishii, D. N. & Efstratiadis, A. (1985) *Nucleic Acids Res.* **13**, 1119–1134
16. Gray, A., Tam, A. W., Dull, T. J., Hayflick, J., Pintar, J., Cavenee, W. K., Koufos, A. & Ullrich, A. (1987) *DNA* **6**, 283–295
17. De Larco, J. E. & Todaro, G. J. (1978) *Nature (London)* **272**, 356–358
18. Blatt, J., White, C., Dienes, S., Friedman, H. & Foley, T. P., Jr. (1984) *Biochem. Biophys. Res. Commun.* **123**, 373–376
19. Reeve, A. E., Eccles, M. R., Wilkins, R. J., Bell, G. I. & Millow, L. J. (1985) *Nature (London)* **317**, 258–260
20. Huff, K. K., Kaufman, D., Gabbay, K. H., Spencer, E. M., Lippman, M. E. & Dickson, R. B. (1986) *Cancer Res.* **46**, 4613–4619
21. Tricoli, J. V., Rall, L. B., Karakousis, C. P., Herrera, L., Petrelli, N. J., Bell, G. I. & Shows, T. B. (1986) *Cancer Res.* **46**, 6169–6173
22. Höppener, J. W. M., Steenbergh, P. H., Slebos, R. J. C., de Pagter-Holthuizen, P., Roos, B. A., Jansen, M., Van den Brande, J. L., Sussenbach, J. S., Jansz, H. S. & Lips, C. J. M. (1987) *FEBS Lett.* **215**, 122–126
23. Sporn, M. B. & Todaro, G. J. (1980) *N. Engl. J. Med.* **303**, 878–880
24. Sporn, M. B. & Roberts, A. B. (1985) *Nature (London)* **313**, 745–747
25. Rotwein, P. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 77–81
26. Jansen, M., Van Schaik, F. M. A., Van Tol, H., Van den Brande, J. L. & Sussenbach, J. S. (1985) *FEBS Lett.* **179**, 243–246
27. Sussenbach, J. S. (1989) *Prog. Growth Factor Res.* **1**, 33–48
28. De Pagter-Holthuizen, P., Jansen, M., Van Schaik, F. M. A., Van der Kammen, R., Oosterwijk, C., Van den Brande, J. L. & Sussenbach, J. S. (1987) *FEBS Lett.* **214**, 259–264
29. Ueno, T., Takahashi, K., Matsuguchi, T., Endo, H. & Yamamoto, M. (1988) *Biochim. Biophys. Acta* **950**, 411–419
30. Ikejiri, K., Ueno, T., Matsuguchi, T., Takahashi, K., Endo, H. & Yamamoto, M. (1990) *Biochim. Biophys. Acta* **1049**, 350–353
31. Yamamoto, M., Maehara, Y., Takahashi, K. & Endo, H. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 7524–7527
32. Thomas, P. S. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 5201–5205
33. Feinberg, A. P. & Vogelstein, B. (1983) *Anal. Biochem.* **132**, 6–13
34. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467
35. Mizusawa, S., Nishimura, S. & Seela, F. (1986) *Nucleic Acids Res.* **14**, 1319–1324
36. Hattori, M. & Sakaki, Y. (1986) *Anal. Biochem.* **152**, 232–238
37. Gubler, U. & Hoffman, B. J. (1983) *Gene* **25**, 263–269
38. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
39. Ueno, T., Takahashi, K., Matsuguchi, T., Endo, H. & Yamamoto, M. (1987) *Biochem. Biophys. Res. Commun.* **148**, 344–349
40. Dull, T. J., Gray, A., Hayflick, J. S. & Ulrich, A. (1984) *Nature (London)* **310**, 777–781
41. Breathnach, R. & Chambon, P. (1981) *Annu. Rev. Biochem.* **50**, 349–383
42. Feng, G., Endo, H. & Yamamoto, M. (1989) *Nucleic Acids Res.* **17**, 4629–4636
43. Padgett, R. A., Grabowski, P. J., Konarska, M. M., Seiler, S. & Sharp, P. A. (1986) *Annu. Rev. Biochem.* **55**, 1119–1150
44. Ueno, T., Takahashi, K., Matsuguchi, T., Ikejiri, K., Endo, H. & Yamamoto, M. (1988) *Carcinogenesis* **9**, 1779–1783
45. Evans, T., DeChiara, T. & Efstratiadis, A. A. (1988) *J. Mol. Biol.* **199**, 61–81
46. Matsuguchi, T., Takahashi, K., Ikejiri, K., Ueno, T., Endo, H. & Yamamoto, M. (1990) *Biochim. Biophys. Acta* **1048**, 165–170
47. Han, V. K. M., D'Ercole, A. J. & Lund, P. K. (1987) *Science* **236**, 193–197
48. DeChiara, T. M., Efstratiadis, A. & Robertson, E. J. (1990) *Nature (London)* **345**, 78–80
49. Yee, D., Cullen, K. J., Paik, S., Perdue, J. F., Hampton, B., Schwartz, A., Lippman, M. E. & Rosen, N. (1988) *Cancer Res.* **48**, 6691–6696

Received 1 July 1991; accepted 12 July 1991