Volume 222, number 1, 181–185 FEB 05157 September 1987

A new 5′-non-coding region for human placental insulin-like growth factor II mRNA expression

Y. Le Bouc, P. Noguiez, P. Sondermeijer*, D. Dreyer*, F. Girard and M. Binoux

INSERM U 142, Hôpital Trousseau, 26 av. Dr Netter, 75012 Paris and **Transgène SA, 11 rue de Molsheim 67000 Strasbourg, France*

Received 17 August 1987

A human placenta cDNA library was screened for insulin-like growth factor II (IGF-II). Four clones were selected, which exhibited an IGF-II cDNA coding sequence identical to those previously described for human adult liver IGF-II cDNA. Extensive sequence diversity was observed in the 5′-non-coding region, probably resulting from differential intron splicing.

Insulin-like growth factor II; Somatomedin; cDNA cloning; Intron splicing; (Human placenta)

1. INTRODUCTION

The insulin-like growth factors (IGF) or somatomedins represent a family of polypeptides which have growth-promoting and insulin-like metabolic effects [1]. cDNA sequences encoding human IGF-I and IGF-II, and part of each IGF gene have been described [2–12]. A single IGF-II gene is present on chromosome 11 close to the insulin gene [10,11]. The variability of the IGF-II mRNA may be attributed mostly to differential gene splicing in the 5′-non-coding region, depending on tissue and stage of development [7–12].

IGF-II is synthesised by various tissues and seems to be predominant during fetal life [13]. We therefore screened a human placenta cDNA library with IGF-II specific probes and analysed the sequence of four cDNAs in detail. When compared with the cDNA sequence described for human adult liver [7,9], differences emerged, particularly in the 5′-non-coding region, but three of the cDNAs had sequences which were homologous with those of the rat liver cell line, BRL 3A [8], or contained a sequence identical to that described for human hepatoma cell, HepG2 [12].

Finally, one of the cDNA clones contained a 5′-non-coding region not previously described.

Correspondence address: Y. Le Bouc, INSERM U 142, Hôpital Trousseau, 26 av. Dr Netter, 75012 Paris, France

The nucleotide sequence presented here has been submitted to the EMBL/GenBank database under the accession number Y00693

2. MATERIALS AND METHODS

2.1. *RNA isolation and cDNA synthesis*

Human term placenta was washed extensively, within 15 min of delivery, in 20 mM Tris-HCl (pH 7.4), 50 mM KCl, 10 mM $MgCl_2$, 100 mM NH_4Cl, 0.5 mM EDTA, and stored at −80°C. RNA was extracted using the guanidinium salt procedure, enriched in poly(A)$^+$ RNA on oligo(dT)-cellulose, then first- and second-strand synthesis and S_1 digestion were performed as in [5].

2.2. *Construction and screening of human cDNA library*

The library was constructed using a synthetic adaptator and screened as described [5]. cDNA was integrated within the one *Eco*RI restriction site

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/87/$3.50

in the C1 gene of λgt10. The library had a complexity of 2×10^6 recombinants. Screening was done by plating 5×10^3 pfu per dish (∅ 130 mm) on a lawn of *E. coli* POP 101. Nitrocellulose filters were hybridised at 50°C overnight in buffer containing 1 M NaCl, 0.1% SDS, 10 mM Tris-HCl (pH 7.5), 0.05% pyrophosphate, 4 × Denhardt's solution (1 × Denhardt's = 0.02% Ficoll, 0.02% bovine serum albumin, 0.02% polyvinyl pyrrolidone), 5% dextran sulfate, and labelled probe (10^6 cpm/ml buffer). The filters were washed at 42°C in 3 × SSC (1 × SSC = 150 mM NaCl, 15 mM Na citrate, pH 7), 0.1% SDS for 3 × 10 min and autoradiographed with intensifying screens.

2.3. *DNA probes*

The human placenta cDNA library was screened for IGF-II with synthetic oligonucleotides complementary to the sequence of human IGF-II mRNA [7,9] encoding amino acids −23 to −17 (probe a, $N = 21$); 56 to 61 (probe b, $N = 17$) and 76 to 82 (probe c, $N = 21$) (fig.2). These probes were 5′-end-labelled using [γ-^{32}P]ATP (3000 Ci/mmol; Amersham, England), and T_4 polynucleotide kinase. We also used the *Bam*HI-*Eco*RI restriction fragment (666 base pairs) of the human adult liver IGF-I cDNA [5] which contains the entire coding sequence. This fragment was labelled by nick translation using [α-^{32}P]dATP (3000 Ci/mmol; Amersham, England).

2.4. *DNA sequence analysis*

cDNA inserts were sequenced using both M13 and IGF-II specific oligonucleotides to prime the dideoxy-chain-terminating reaction. The strategy followed for each of the clones is outlined in fig.1. Oligonucleotide d was complementary to nucleotide sequence-encoding amino acids 119–123 (fig.2B).

3. RESULTS

From the initial 160000 recombinants screened with probe c, six were selected for further hybridisation. Two of them (P35, P44) gave a signal with oligonucleotide b (P35) or cross-hybridised with a nick-translated IGF-I cDNA probe (P35, P44; fig.1). The sequence of the *Eco*RI inserts for P35 and P44 (663 and 123 bp, respectively) is shown in fig.2. The region coding

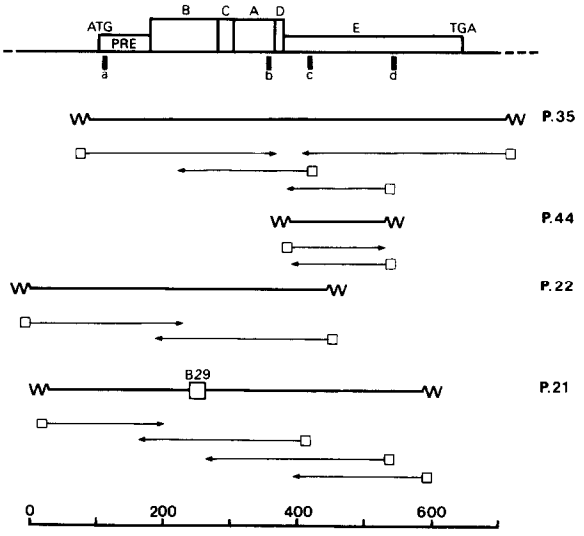

Fig.1. Schematic representation of the four human placental cDNA inserts encoding IGF-II (clones P22, P35, P44) and an IGF-II variant (clone P21). Regions coding for the signal peptide (PRE), and the B, C, A, D and E domains are depicted by open boxes, the 5′- and 3′-untranslated regions being indicated by thin lines. The positions of the synthetic oligonucleotides (a–d) complementary to the nucleotide sequence of mRNA are indicated. The strategy for DNA sequence determination is indicated below each of the clones sequenced. Zigzag lines symbolise M13 arms, open squares represent primers (universal sequencing primer, oligonucleotide a, c and d) used for chain elongation, and the horizontal arrows indicate the direction and extent of sequence determinations. B29 signifies the nucleotide sequence variation in the IGF-II variant cDNA described by Jansen et al. [9].

for the precursor of IGF-II was identical to that described in [7,9]. However, the sequence corresponding to the 5′-non-coding exon, which ends seven nucleotides before the ATG initiator at amino acid position −24, varied considerably when different IGF-II cDNAs were compared (fig.2). For instance, this part of the sequence in P35 did not correspond to the structure published for human adult liver [7,9] but matched the sequence described for the rat liver cell line, BRL 3A [8,14,15].

A second screening of 190000 recombinants from the same library was carried out with an oligonucleotide closer to the 5′-end in the cDNA sequence (probe a, fig.1). The structure of the two

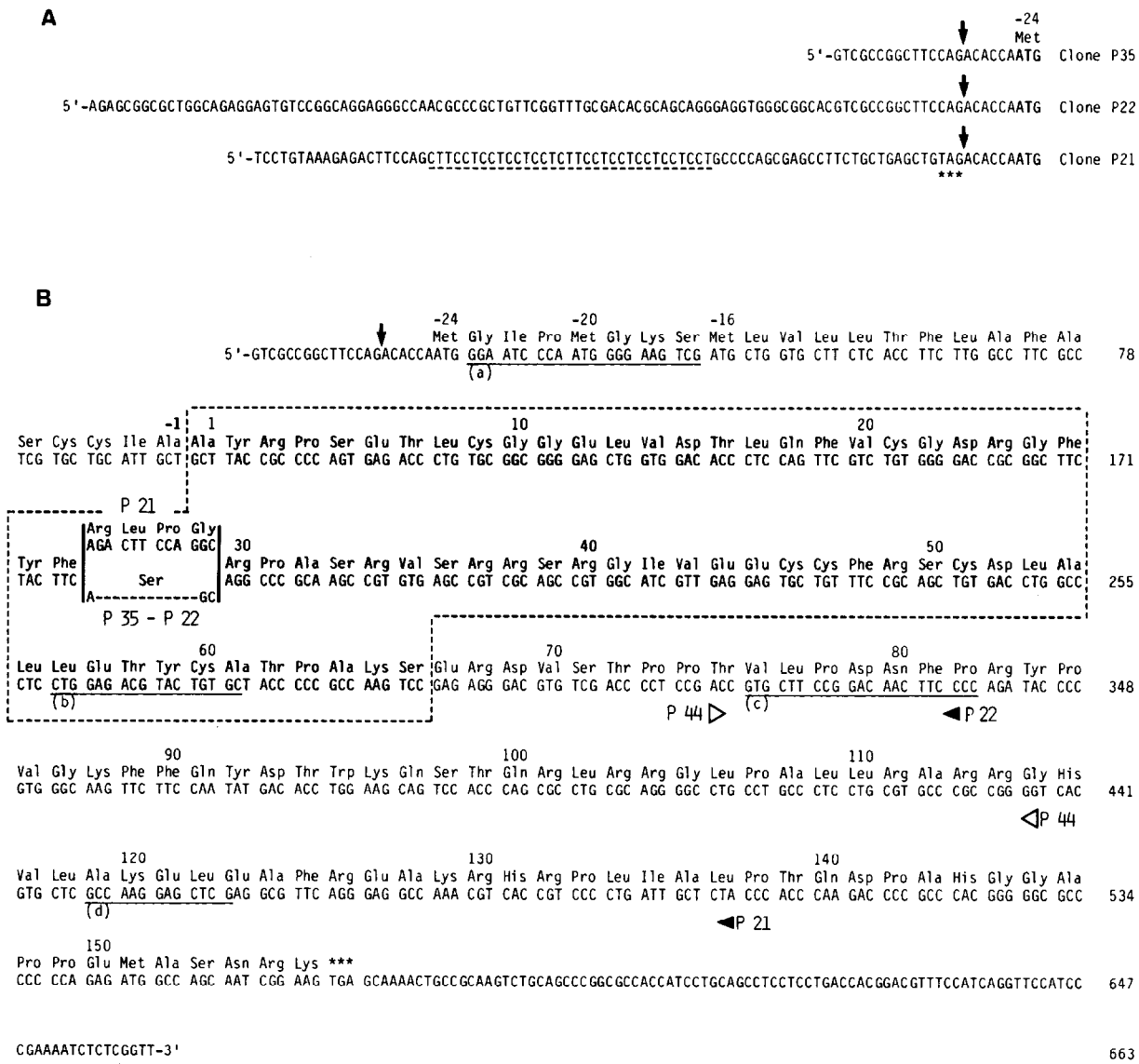

Fig.2. Nucleotide sequence of the human placental cDNA inserts encoding IGF-II and an IGF-II variant. (A) Comparison of the nucleotide sequences of the 5'-untranslated regions of the three IGF-II cDNA clones (P35, P22, P21). The sequences were identical over six nucleotides after the initiator ATG, then diverged after the splice site (indicated by vertical arrows). The simple trinucleotide repeat (P21) is underlined (dashed line). Asterisks indicate the stop codons, and Met −24 the ATG codon at which translation starts. (B) Nucleotide and deduced amino acid sequences of IGF-II cDNA insert P35. The cDNA strand presented is homologous to the IGF-II mRNA. The number of the nucleotide at the end of each line is indicated. The amino acid sequence shown is numbered from the amino-terminus of the IGF-II molecule. The amino acids preceding mature IGF-II are shown as negative numbers. The region corresponding to the 67 amino acid residues of IGF-II is boxed. The asterisks indicate the termination codon (TGA). The underlined nucleotides designate the sequence complementary to synthetic oligonucleotides (a–d) used as probes or primers. The sequence of clone P44 is indicated between ▷ and ◁. The 3'-end of the P21 and P22 cDNA inserts is indicated by ◀. Their 5'-end is shown in panel A. In the B domain, the sequences of P22–P35 and P21 diverge; the amino acids and nucleotides which correspond are bracketed.

isolates, P21 and P22, with insert sizes of 580 and 423 bp, respectively, is described in fig.2. P22 had a coding region identical to that of P35. The sequence corresponding to the first exon, extending 107 nucleotides upstream of the ATG initiator, confirmed the homology with the IGF-II cDNA described for the rat liver cell line, BRL 3A [8,14,15], as already observed in the case of P35. However, a different sequence was observed in the first exon of P21. This region resembled none of the previously described IGF-II cDNAs and contained a repeated CCT trinucleotide 35 nucleotides upstream of the ATG initiator. The coding region of P21 corresponded to the B29 IGF-II variant cDNA [9], which uses an alternative splice site between the first two coding exons.

4. DISCUSSION

The coding sequence of human placental IGF-II RNA is identical to that described previously for the messenger isolated from human liver [7,9]. However, the sequence of the 5′-non-coding region ending seven nucleotides before the ATG initiator exhibits extensive variability, not only from one tissue to another, but also within the messenger population of a particular organ, such as the liver or placenta. The 5′-end of some of these clones was homologous with one of those found in rat liver cell line (BRL 3A) cDNA [8,14,15]. This exon, which has been mapped in the IGF-II gene region of chromosome 11 [8,11,14], could be transcribed not only in human placental, but also in human hepatoma cell line HepG2, IGF-II cDNA [12].

In addition, one clone (P21) contained a sequence in this region which has not been described before. A structure of repeated trinucleotides was found at a distance of 35 nucleotides upstream of the start of the coding region within the IGF-II cDNA. This sequence is not derived from the intron preceding the first coding exon [11], which would have indicated incomplete processing of the primary transcript. This suggests that an exon exists in addition to those already described [7–12,14]. Regions of simple repetitive DNA sequences in the vicinity of structural genes have been found in other gene families, whether in the 5′-flanking region [16–19], intervening sequences [20–22] or satellite DNA [23]. It remains to be established whether these short CCT repeats are in any way related to the regulation of IGF-II gene expression. The coding sequence of P21 clone with this simple DNA repeat contains the B29 IGF-II variant (fig.2) described by Jansen et al. [9]. These results show that the selection of the 5′-non-coding region is regulated independently of the alternative splicing of B29 IGF-II variant mRNA.

The IGF-II gene, which is present as a single copy on the human genome [10,11], is generally transcribed in messengers with an identical coding region but with different 5′-terminal non-coding sequences [7–9,11,12,14,15]. It would seem that, in eucaryotic systems, transcription must be initiated from different sites [24–26] or else the primary transcript is processed in more than one way [27]. Recent findings concerning the IGF-II gene indicate that both mechanisms occur. In the case of the rat, two types of promoter region have been described [14,15], and the promoter used in adult human tissues has been shown to differ from that used in fetal tissue [12]. An alternative splicing mechanism generating a different mature messenger has been described for a cDNA variant of IGF-II [9,11]. The structure of the 5′-non-coding region described here for the clone, P21, may result from additional promoter regions upstream in the gene, or from variations in the processing of the transcript. Structural domains of a given protein are often encoded by individual exons [19] and to this extent one can only speculate as to the reasons for heterogeneity in the 5′-non-coding sequence of a single messenger. One possibility might be that it reflects the use of multiple promoters as a mode of regulating gene activity [14,15,24]. Alternatively, the stability [28,29] or translation efficiency [27,30] of the mature messenger may be affected by this heterogeneity. In the case of the IGF-II gene, initiation of transcription and processing of the primary transcript are apparently not regulated by single, clearly defined mechanisms, as is usually observed for eucaryotic genes.

The biological significance of this phenomenon remains an open and intriguing question.

ACKNOWLEDGEMENTS

This work was supported by the Institut National de la Santé et de la Recherche Médicale (IN-

SERM) and in part by a grant from Nordisk Insulinlaboratorium (Denmark) (to P.N.).

REFERENCES

[1] Zapf, J., Schmid, C. and Froesch, E.R. (1984) Clin. Endocrinol. Metab. 13, 3–30.
[2] Jansen, M., Van Schaik, F.M.A., Ricker, A.T., Bullock, B., Woods, D.E., Gabbay, K.H., Nussbaum, A.L., Sussenbach, J.S. and Van den Brande, J.L. (1983) Nature 306, 609–611.
[3] Ullrich, A., Berman, C.H., Dull, T.J., Gray, A. and Lee, J.M. (1984) EMBO J. 3, 361–364.
[4] Rotwein, P. (1986) Proc. Natl. Acad. Sci. USA 83, 77–81.
[5] Le Bouc, Y., Dreyer, D., Jaeger, F., Binoux, M. and Sondermeyer, P. (1986) FEBS Lett. 196, 108–112.
[6] Rotwein, P., Pollock, K.M., Didier, D.K. and Krivi, G.G. (1986) J. Biol. Chem. 261, 4828–4832.
[7] Bell, G.I., Merryweather, J.P., Sanchez-Pescador, R., Stempiem, M.M., Priestley, L., Scott, J. and Rall, L.B. (1984) Nature 310, 775–777.
[8] Dull, T.J., Gray, A., Hayflick, J.S. and Ullrich, A. (1984) Nature 310, 777–781.
[9] Jansen, M., Van Schaik, F.M.A., Van Tol, H., Van den Brande, J.L. and Sussenbach, J.S. (1985) FEBS Lett. 179, 243–246.
[10] Bell, G.I., Gerhard, D.S., Fong, N.M., Sanchez-Pescador, R. and Rall, L.B. (1985) Proc. Natl. Acad. Sci. USA 82, 6450–6454.
[11] De Pagter-Holthuizen, P., Van Schaik, F.M.A., Verduijn, G.M., Van Ommen, G.J.B., Bouma, B.N., Jansen, M. and Sussenbach, J.S. (1986) FEBS Lett. 195, 179–184.
[12] De Pagter-Holthuizen, P., Jansen, M., Van Schaik, F.M.A., Van der Kammen, R., Oosterwijk, C., Van den Brande, J.L. and Sussenbach, J.S. (1987) FEBS Lett. 214, 259–264.
[13] Nissley, S.P. and Rechler, M.M. (1984) in: Hormonal Proteins and Peptides (Li, C.H. ed.) vol.12, pp.128–195, Academic Press, New York.
[14] 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.
[15] Frunzio, R., Chiariotti, L., Brown, A.L., Graham, D.E., Rechler, M.M. and Bruni, C.B. (1986) J. Biol. Chem. 261, 17138–17149.
[16] Bender, T.P. and Kuehl, W.M. (1986) Proc. Natl. Acad. Sci. USA 83, 3204–3208.
[17] Cohen, J.B., Effron, K., Rechavi, G., Ben-Neriah, Y., Zakut, R. and Givol, D. (1982) Nucleic Acids Res. 11, 3353–3370.
[18] Edens, L., Heslinga, L., Klok, R., Ledeboer, A.M., Maat, J., Toonen, M.Y., Visser, C. and Verrips, C.T. (1982) Gene 18, 1–12.
[19] Nef, P., Mauron, A., Stalder, R., Alliod, C. and Ballivet, M. (1984) Proc. Natl. Acad. Sci. USA 81, 7975–7979.
[20] Moormann, R.J.M., Den Dunnen, J.T., Mulleners, L., Andreoli, P., Bloemendal, H. and Schoenmakers, J.G.G. (1983) J. Mol. Biol. 171, 353–368.
[21] Hall, J.L. and Cowan, N.J. (1985) Nucleic Acids Res. 13, 207–223.
[22] Qasba, P.K. and Safaya, S.K. (1984) Nature 308, 377–380.
[23] Fowler, R.F., Bonnewell, V., Spann, M.S. and Skinner, D.M. (1985) J. Biol. Chem. 260, 8964–8972.
[24] Schibler, U., Hagenbuchle, O., Wellauer, P.K. and Pittet, A.C. (1983) Cell 33, 501–508.
[25] Benyajati, C., Spoerel, N., Haymerle, H. and Ashburner, M. (1983) Cell 33, 125–133.
[26] Carlson, M. and Botstein, D. (1982) Cell 28, 145–154.
[27] Reynolds, G.A., Goldstein, J.L. and Brown, M.S. (1985) J. Biol. Chem. 260, 10369–10377.
[28] Bentley, D.L. and Groudine, M. (1986) Mol. Cell. Biol. 10, 3481–3489.
[29] Nishikura, K. (1987) J. Mol. Biol. 193, 497–505.
[30] McGarry, T.J. and Lindquist, S. (1985) Cell 42, 903–911.