Organization of the human genes for insulin-like growth factors I and II

P. de Pagter-Holthuizen, F.M.A. van Schaik, G.M. Verduijn, G.J.B. van Ommen*, B.N. Bouma°, M. Jansen+ and J.S. Sussenbach

Laboratory for Physiologieal Chemistry, State University of Utrecht, Vondellaan 24A, 3521 GG Utrecht, °*Dept of Haematology,* +*Dept of Pediatrics, State University of Utrecht, Utrecht* and **Dept of Human Genetics, State University of Leiden, Leiden, The Netherlands*

Received 9 October 1985

Recently, we have reported the isolation of cDNAs encoding the precursors of insulin-like growth factors I and II (IGF-I and II) [(1983) Nature 306, 609–611; (1985) FEBS Lett. 179, 243–246. These cDNAs were employed as specific probes to detect and isolate the corresponding genes from human cosmid DNA libraries. Three cosmids were detected, together containing the entire cDNA sequence of IGF-I, and one cosmid containing the sequence of IGF-II cDNA. Southern blot hybridization, physical mapping and nucleotide sequence analysis of these cosmids revealed that the IGF-I and -II genes have a discontinous structure. The IGF-I gene contains at least four exons spanning a region of probably more that 45 kilobasepairs (kb), while the IGF-II gene consists of at least five exons, spanning a region of 16 kb.

Insulin-like growth factor I *Insulin-like growth factor II* *Somatomedin* *Gene structure* *Hormone precursor*

1. INTRODUCTION

The insulin-like growth factors (IGF) constitute a heterogeneous group of hormonal polypeptides with insulin-like and growth-promoting properties. The amino acid sequence of 2 of the human IGFs have been identified [3,4]. IGF-I consists of 70 amino acids, while IGF-II, which is composed of 67 amino acids, shares homology with the so-called multiplication stimulating activity (MSA) in the rat [5]. Both IGF-I and IGF-II show a high degree of homology.

We have reported the sequences of cDNAs encoding human IGF-I and IGF-II [1,2]. These sequences reveal that the coding regions of IGF-I and IGF-II are flanked by regions encoding aminoterminal as well as carboxyl-terminal peptides, indicating that both growth factors are synthesized as precursor molecules. The nucleotide sequence of human IGF-II cDNA was also determined by Bell et al. [6].

Using the cDNAs as specific probes, the chromosomal assignment has been determined. The IGF-I gene maps to chromosome 12, whereas the gene for IGF-II is located on chromosome 11 [7–10]. Here, we report the genomic organization of the coding and the flanking non-coding regions of the human IGF-I and IGF-II genes.

2. MATERIALS AND METHODS

2.1. *Genomic libraries*

Two non-amplified human genomic cosmid libraries were constructed from human placenta DNA as described [11]. A third cosmid genomic library was constructed from GM 1416, 48 XXXX cell-line DNA, using a double-cos-site vector 2cRB [12]. Cosmids were transduced into *E. coli* 1046 and of each library 200000 colonies were screened basically as described by Benton and Davis [13]. Nitrocellulose filter (90 mm, Satorius), containing 20000 colonies each, were washed after hybridiza-

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/86/$3.50 © 1986 Federation of European Biochemical Societies

tion under stringent conditions (0.3 × SSC, 0.1% SDS, pH 7.4) at 65°C.

2.2. *DNA probes*

cDNA probes were isolated from the plasmids pIGF-I, pIGF-II and pIGF-II var [1,2]. The 731 bp cDNA probe of IGF-I consists of the coding region of the IGF-I precursor, flanked by the 5′- and 3′-non-translated regions of 80 and 264 bp, respectively. As probes for IGF-II we used partially overlapping cDNAs containing the coding regions of the IGF-II precursor, flanked by 5′- and 3′-non-translated regions of 488 and 237 bp, respectively. Double-stranded cDNA probes were labeled by nick-translation with [α-^{32}P]dCTP to a specificity of 10^8 cpm/μg DNA.

```
                   b
                   ▽
...TCTGTGT  TTTGTAGATA  AATGTGAGGA  TTTTCTCTAA  ATCCCTCTTC  TGTTTGCTAA  ATCTCACTGT  CACTGCTAAA  TTCAGAGCAG  ATAGAGCCCG

-1
CGCAATGGAA  TAAAGTCCTC  AAAATTGAAA  TGTGACATTG  CTCTCAACAT  CTCCCATCTC  TCTGGATTCT  TTTGCTTCAT  TATTCCTGCT  AACCAATTCA
       a
       ▽        -80         -70         -60         -50         -40         -30         -20         -10  ▼
CTTTCAGACT  TTGTACTTCA  GAAGCAATGG  GAAAAATCAG  CAGTCTTCCA  ACCCAATTAT  TTAAGTGCTG  CTTTTGTGAT  TTCTTGAAGG  TAAATATTTC

TTACTCTT..


1
                   ▼                    10          20          30              40          50          60
... TTT CTT GGG CAG GTG AAG ATG CAC ACC ATG TCC TCC TCG CAT CTC TTC TAC CTG GCG CTG TGC CTG CTC ACC TTC ACC AGC TCT GCC
                            Met His Thr Met Ser Ser Ser His Leu Phe Tyr Leu Ala Leu Cys Leu Leu Thr Phe Thr Ser Ser Ala

70          80          90              100         110         120         130         140         150▼
ACG GCT GGA CCG GAG ACG CTC TGC GGG GCT GAG CTG GTG GAT GCT CTT CAG TTC GTG TGT GGA GAC AGG GGC TTT TAT TTC AGT AAG T..
Thr Ala Gly Pro Glu Thr Leu Cys Gly Ala Glu Leu Val Asp Ala Leu Gln Phe Val Cys Gly Asp Arg Gly Phe Tyr Phe A


2
                    ▼           160         170         180             190         200         210             220
... TTG ATT TGT GCA GAC AAG CCC ACA GGG TAT GGC TCC AGC AGT CGG AGG GCG CCT CAG ACA GGT ATC GTG GAT GAG TGC TGC TTC CGG
                     sn Lys Pro Thr Gly Tyr Gly Ser Ser Ser Arg Arg Ala Pro Gln Thr Gly Ile Val Asp Glu Cys Cys Phe Arg

    230             240         250         260         270             280         290         300             310
AGC TGT GAT CTA AGG AGG CTG GAG ATG TAT TGC GCA CCC CTC AAG CCT GCC AAG TCA GCT CGC TCT GTC CGT GCC CAG CGC CAC ACC GAC
Ser Cys Asp Leu Arg Arg Leu Glu Met Tyr Cys Ala Pro Leu Lys Pro Ala Lys Ser Ala Arg Ser Val Arg Ala Gln Arg His Thr Asp

    320         330    ▼
ATG CCC AAG ACC CAG AAG GTA AGC CCA CCT ...
Met Pro Lys Thr Gln Lys


3
                   ▼        340         350         360             370         380         390             400
... CTT TAT TTA TAG GAA GTA CAT TTG AAG AAC GCA AGT AGA GGG AGT GCA GGA AAC AAG AAC TAC AGG ATG TAG GAA GAC CCT CCT GAG
                    Glu Val His Leu Lys Asn Ala Ser Arg Gly Ser Ala Gly Asn Lys Asn Tyr Arg Met ---

410         420             430         440         450             460         470         480             490
GAG TGA AGA GTG ACA TGC CAC CGC AGG ATC CTT TGC TCT GCA CGA GTT ACC TGT TAA ACT TTG GAA CAC CTA CCA AAA AAT AAG TTT GAT

500         510             520         530         540             550         560         570             580
AAC ATT TAA AAG ATG GGC GTT TCC CCC AAT GAA ATA CAC AAG TAA ACA TTC CAA CAT TGT CTT TAG GAG TGA TTT GCA CCT TGC AAA AAT

590         600             610         620         630             640         650
GGT CCT GGA GTT GGT AGA TTG CTG TTG ATC TTT TAT CAA TAA TGT TCT ATA GAA AAG AAA AAA AAC ACA CAC ACA CAC ACA CAC TTA GTC

CTG TCC TAG ... ...
```

Fig.1. Nucleotide sequence of the human IGF-I gene. The numbers −1, 1, 2 and 3 refer to the corresponding exons. The exon-intron boundaries are indicated by ▼; $\overset{a}{\triangledown}$ and $\overset{b}{\triangledown}$ are the 2 possible 5′-splice sites of exon −1. Starting with the ATG codon for initiation of translation the sequence corresponding to the cDNA is numbered positive in the 5′⟶3′ direction and negative in the 3′⟶5′ direction. The prepro-IGF-I amino acid sequence is shown, in which the underlined part indicates the mature IGF-I peptide.

2.3. *Southern blotting of cosmid DNA*

Cosmid DNA was isolated according to standard procedures [14]. 1 μg cosmid DNA was digested with several restriction enzymes, run on an 0.7% agarose gel and transferred to nitrocellulose (Schleicher and Schüll). Hybridizations were performed in the presence of 50% formamide at 42°C (18 h). Blots were washed under stringent conditions at 65°C, 1 × SSC, 0.1% SDS, pH 7.4.

2.4. *Restriction mapping, subcloning and nucleotide sequence analysis*

The cosmid clones containing parts of the human IGF-I or IGF-II genes were purified. After restriction enzyme analysis, fragments carrying the human IGF-I or IGF-II sequences were identified by Southern blot hybridization. Using [α-^{32}P] dCTP specific cDNA probes restriction maps of these cosmids were constructed. Some cosmid fragments were isolated directly, other fragments were first subcloned in pUC12 followed by sequence analysis. Nucleotide sequence analysis was carried out according to Maxam and Gilbert [15].

3. RESULTS

Employing ^{32}P-labeled cDNA probes encoding IGF-I and IGF-II, respectively, 3 human genomic cosmid DNA libraries were screened for IGF-I and -II specific clones as described in section 2. Three different clones (clones a, b and c) hybridized specifically with IGF-I cDNA, whereas only a single clone showing homology to IGF-II cDNA was isolated (cos IGF-II).

Southern blot hybridization analysis of *Hin*dIII-digested total human placenta DNA and the above-mentioned cosmid DNAs revealed that the entire IGF-I and IGF-II cDNA sequences were present in these 4 cosmids (not shown). Employing detailed restriction enzyme analysis, Southern blot hybridization and nucleotide sequence analysis of the cosmids, we have determined the precise position of the IGF-I and IGF-II specific sequences on the physical maps of the cosmids described above (figs 1–3). This analysis reveals that the genes coding for IGF-I and IGF-II have a discontinuous structure. The 713 bp IGF-I cDNA is encoded by 4 exons extended over at least 45 kb of genomic DNA, while the 1265 bp IGF-II cDNA sequence is encoded by 5 exons spanning a region of 16 kb of chromosomal DNA.

4. DISCUSSION

We have studied the organization of the genes coding for human IGF-I and IGF-II. Employing Southern blot hybridization and nucleotide sequence analysis the IGF-I and IGF-II cDNAs containing the entire coding sequence for the precursor as well as the 5′- and 3′-non-coding sequences have been positioned on genomic cosmid clones.

Comparison of the structure of the exons of the IGF-I and IGF-II genes shows some interesting common features. Both genes contain an exon (exon 1) which starts 6 bp in front of the potential ATG initiation codon and terminates close to the COOH-terminus of the B-domain. Also, both genes contain an exon starting at the end of the B-domain and terminating in the E peptide of the growth factor precursors (exon 2), while in both genes the COOH-termini of the precursor are located in a single exon (exon 3, fig.3). On the other hand, the size of the corresponding introns seems to differ extensively. The gene for IGF-I must be of considerable length since the 713 bp cDNA is divided over 4 exons which are located on 3 non-overlapping cosmid DNAs, suggesting that these exons are located on a stretch of human chromosomal DNA of at least 45 kb. Although our cosmid clones contain the entire coding region of the IGF-I precursor they probably do not represent the entire IGF-I mRNA, because the cDNA appears to be incomplete. The previously assigned poly(A) tail [1] is also present in genomic DNA indicating that this A-rich region in IGF-I cDNA does not represent the actual poly(A) tail.

The gene for human IGF-II spans a region of at least 16 kb chromosomal DNA. The positions of exons 1, 2 and 3 have also been determined by Dull et al. [17], who used a rat MS cDNA as probe. As described in [2], 2 closely related but distinct types of IGF-II cDNA were isolated from the same human liver cDNA library. The difference of the IGF-II variant (IGF-II var) compared to IGF-II is situated at amino acid 29, where the variant IGF-II cDNA predicts a tetrapeptide Arg-Leu-Pro-Gly due to an insert of 9 nucleotides at position 158 (fig.2). Interestingly, examination of the genomic

-2

```
                         -460       -450       -440       -430       -420       -410       -400
...CTTCGTT TCTCCAGCCT CAGCTTTGTC CCTCTCCTCC TCCACGTCAA CCTGGCCAGA GGGTCTGGAC GCCACAGCCA GGGCACCCCC
   -390       -380       -370       -360       -350       -340       -330       -320       -310
TGCTTTGGTG GTGACTGCTA ATATTGGCCA GGCCGGCGGA TCATCGTCCA GGCAGTTTCG GCAGAGAGCC TTGGGCACCA GTGACTCCCC
   -300       -290       -280       -270       -260       -250
GGTCCTCTTT ATCCACTGTC CAGGAGCTGC GGGGACTGCG CAGGGACTAG AGTACAGGTA ACTGGGCTCC CAT.......
```

-1

```
                         -240       -230       -220       -210       -200       -190       -180
...TTTGTTT TTCCCAGGGG CCGAAGAGTC ACCACCGAGC CTGTGTGGGA GGAGGTGGAT TCCAGCCCCC AGCCCCAGGG CTCTGAATCG
   -170       -160       -150       -140       -130       -120       -110       -100        -90
CTGCCAGCTC AGCCCCCTGC CCAGCCTGCC CCACAGCCTG AGCCCCGGCA GGCCAGAGAG CCCAGTCCTG AGGTGAGCTG CTGTGGCCTG
   -80        -70        -60        -50        -40        -30        -20        -10
TGGCCCAGGC GACCCCAGCG CTCLCAGAAC TGAGGCTGGC AGCCAGCCCC AGCCTCAGCC CCAACTGCGA GGCAGAGAGG TGAGTGTCTC
AGGCA.....
```

1

```
                                        10          20          30              40          50          60
... TCC CGC CCC CAG ACA CCA ATG GGA ATC CCA ATG GGG AAG TCG ATG CTG GTG CTT CTC ACC TTC TTG GCC TTC GCC TCG TGC TGC ATT
                            Met Gly Ile Pro Met Gly Lys Ser Met Leu Val Leu Leu Thr Phe Leu Ala Phe Ala Ser Cys Cys Ile
70          80          90              100         110         120             130         140         150
GCT GCT TAC CGC CCC AGT GAG ACC CTG TGC GGC GGG GAG CTG GTG GAC ACC CTC CAG TTC GTC TGT GGG GAC CGC GGC TTC TAC TTC AGT
Ala Ala Tyr Arg Pro Ser Glu Thr Leu Cys Gly Gly Glu Leu Val Asp Thr Leu Gln Phe Val Cys Gly Asp Arg Gly Phe Tyr Phe S
                                                                                                                     A
AAG TAG CTG GGA GGG ... ...
```

2

```
                                    160         170         180             190         200         210             220
... CTC TGT GCT GTG GGA CTT CCA GGC AGG CCC GCA AGC CGT GTG AGC CGT CGC AGC CGT GGC ATC GTT GAG GAG TGC TGT TTC CGC AGC
                                 er Arg Pro Ala Ser Arg Val Ser Arg Arg Ser Arg Gly Ile Val Glu Glu Cys Cys Phe Arg Ser
                     rg Leu Pro Gly
        230         240             250         260         270             280         290         300
TGT GAC CTG GCC CTC CTG GAG ACG TAC TGT GCT ACC CCC GCC AAG TCC GAG AGG GAC GTG TCG ACC CCT CCG ACC GTG CTT CCG GTG AGG
Cys Asp Leu Ala Leu Leu Glu Thr Tyr Cys Ala Thr Pro Ala Lys Ser Glu Arg Asp Val Ser Thr Pro Pro Thr Val Leu Pro
GTC CTG GGC ... ...
```

3

```
                                310         320         330             340         350         360         370
... ... CCC TTC CCC TCC CAG GAC AAC TTC CCC AGA TAC CCC GTG GGC AAG TTC TTC CAA TAT GAC ACC TGG AAG CAG TCC ACC CAG CGC
                            Asp Asn Phe Pro Arg Tyr Pro Val Gly Lys Phe Phe Gln Tyr Asp Thr Trp Lys Gln Ser Thr Gln Arg
    380         390             400         410         420             430         440         450         460
CTG CGC AGG GGC CTG CCT GCC CTC CTG CGT GCC CGC CGG GGT CAC GTG CTC GCC AAG GAG CTC GAG GCG TTC AGG GAG GCC AAA CGT CAC
Leu Arg Arg Gly Leu Pro Ala Leu Leu Arg Ala Arg Arg Gly His Val Leu Ala Lys Glu Leu Glu Ala Phe Arg Glu Ala Lys Arg His
    470         480             490         500         510             520         530         540         550
CGT CCC CTG ATT GCT CTA CCC ACC CAA GAC CCC GCC CAC GGG GGC GCC CCC CCA GAG ATG GCC AGC AAT CGG AAG TGA GCA AAA CTG CCG
Arg Pro Leu Ile Ala Leu Pro Thr Gln Asp Pro Ala His Gly Gly Ala Pro Pro Glu Met Ala Ser Asn Arg Lys ---
    560         570             580         590         600             610         620         630         640
CAA GTC TGC AGC CCG GCG CCA CCA TCC TGC AGC CTC CTC CTG ACC ACG GAC GTT TCC ATC AGG TTC CAT CCC GAA AAT CTC TCG GTT CCA
    650         660             670         680         690             700         710         720         730
CGT CCC CTG GGG CTT CTC CTG ACC CAG TCC CCG TGC CCC GCC TCC CCG AAA CAG GCT ACT CTC CTC GGC CCC CTC CAT CGG GCT GAG GAA
    740         750             760         770         780
GCA CAG CAG CAT CTT CAA ACA TGT ACA AAA TCG ATT GGC TTT AAA CAC CCT TCA CAT ACC CTC CCC CCA AAT TAT CCC CAA TTA TCC CCA

CAC ATA AAA AAT CAA AAC ATT AAA CTA ACC CCC TTC CCC CCC CCC CAC AAC AAC CCT CTT AAA ACT AAT TGG CTT TTT AGA AAC ACC CCA

CAA AAG CTC AGA AAT TGG CTT TAA AAA AAA CAA CCA CCA AAA AAA ATC AAT TGG CTA AAA AAA AAG CAC CAA AAA CGA CCG GCT GAG AAC

AAT CGC ... ...
```

←

Fig.2. Nucleotide sequence of the human IGF-II gene. The numbers −2, −1, 1, 2 and 3 refer to the corresponding exons. The 5′- and 3′-exon-intron boundaries are indicated by (▼). Starting with the ATG codon for initiation of translation the sequence corresponding to the cDNA is numbered positive in the 5′ ⟶ 3′ direction and negative in the 3′ ⟶ 5′ direction. The prepro-IGF-II amino acid sequence is shown as well as the 3 amino acid insert for IGF-II var. The underlined amino acid sequence indicates the mature IGF-II peptide. A possible alternative poly(A) site ATTAAA is also underlined.

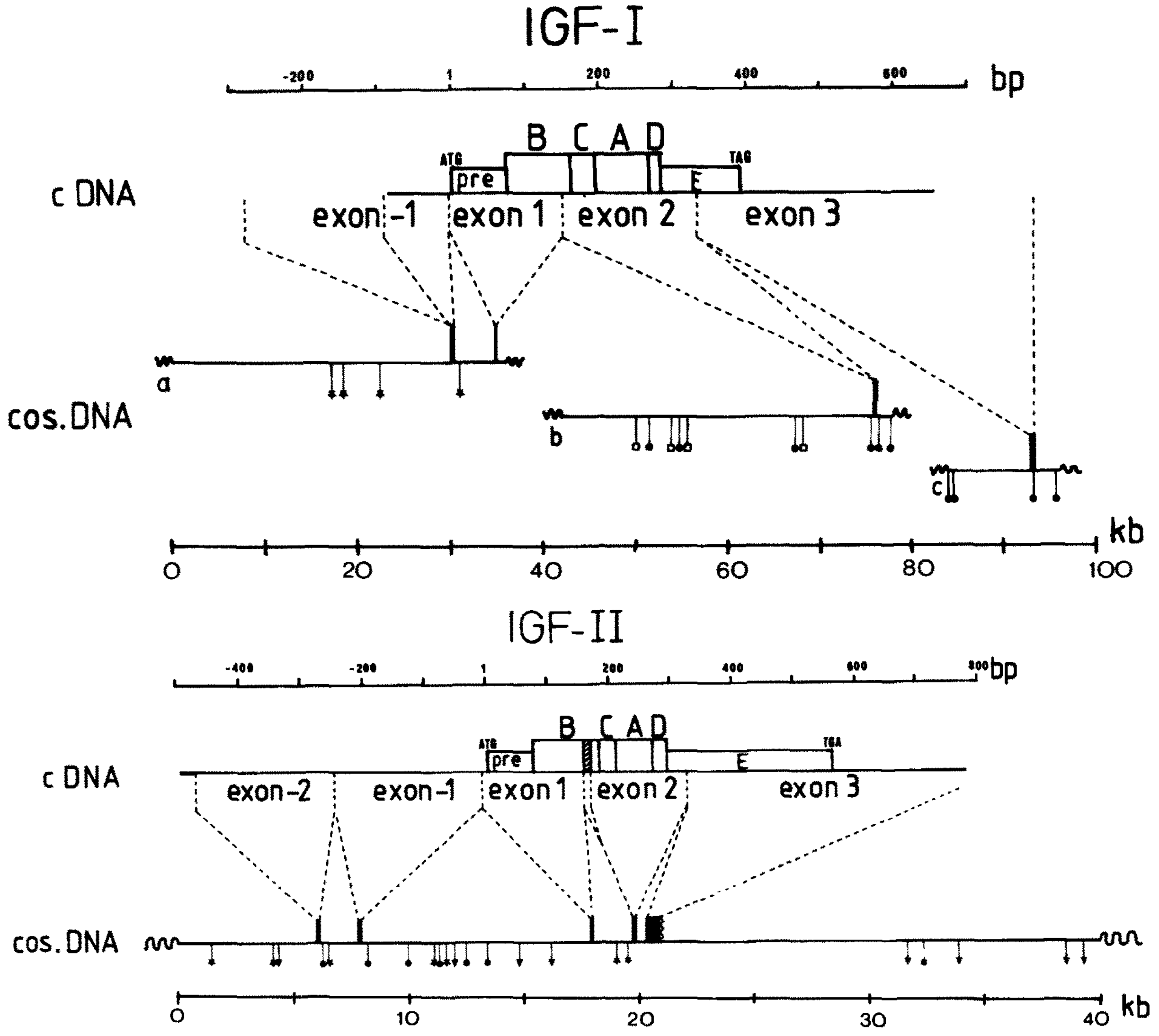

Fig.3. Schematic representation of the IGF-I and IGF-II genes. The cDNA structures containing regions coding for the B, C, A and D domains (higher open boxes), the N-terminal peptide (pre), the C-terminal peptide (E) and the 5′- and 3′-untranslated sequences are schematically depicted. The restriction maps of the genomic cosmid clones, a, b and c for IGF-I and cosIGF-II and the localization of the exons are indicated. The exons are shown by the black boxes numbered negative for the 5′-non-coding regions and positive for the coding regions. The exon carrying the ATG initiation codon of the signal peptide is designated exon 1. Starting with the ATG codon for initiation of translation, the sequence corresponding to the cDNA is numbered positive in the 5′ ⟶ 3′ direction and negative in the 3′ ⟶ 5′ direction. The 5′-end of exon −1 of the IGF-I gene is indicated by 2 potential acceptor splice sites ($\overset{a}{\triangledown}$ and $\overset{b}{\triangledown}$, fig.1). The nonanucleotide insert in the B-domain of IGF-II var cDNA is indicated by the shaded block. Restriction sites: *, *Hind*III; ▼, *Eco*RI; ●, *Bam*HI; □, *Kpn*I.

DNA sequence reveals that the difference is located at the 5′-intron-exon boundary of exon 2. At this position IGF-II var contains a nonanucleotide insertion which is identical to the last 9 bp of the intron for IGF-II cDNA (fig.2, exon 2). This suggests that the corresponding IGF-II var mRNA arises by splicing 9 nucleotides upstream from the acceptor splice junction of IGF-II (fig.2, exon 2). However, the nucleotide sequence at the hypothetical IGF-II var intron-exon junction of exon 2 (CTGTGG) deviates from the canonical acceptor sequence PyPyNPyAG [18]. Furthermore, Atweh et al. [19] have demonstrated that the sequence CCACGG is not functional in splicing. At the moment we favour the explanation that the IGF-II var gene represents an allelic variant of the IGF-II gene containing the acceptor sequence CTGTAG which is present in the human population in low abundance. Interestingly, the IGF-II var gene is expressed in vivo, since recently Humbel has identified its gene product as a minor fraction in pooled sera (Humbel, personal communication). Also, for IGF-II it is unlikely that the cDNAs obtained so far represent the complete messenger RNA, since a polyadenylation signal and a poly(A) tail are absent in the cDNAs. In this respect it is noteworthy that Scott et al. [20] recently demonstrated by Northern blot analysis of total RNA from adult human liver that IGF-II mRNA has a length of 5.3 kb. Further studies of the genomic cosmic clones and isolation of longer cDNAs will be necessary to determine additional 5′- and 3′-non-translated exons.

In conclusion, we have identified at the nucleotide level those parts of the IGF-I and IGF-II genes which correspond to the coding sequences of prepro-IGF-I and -II. The results obtained so far suggest that the IGF-I and -II genes contain long non-coding sequences, spanning an extensive region of the human genome.

ACKNOWLEDGEMENTS

We thank Dr J.L. Van den Brande for stimulating discussions and critical reading of the manuscript, J. van der Rijst for preparation of the figures and Mrs E. Simon for typing the manuscript.

REFERENCES

[1] 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.

[1] 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.

[3] Rinderknecht, E. and Humbel, R.E. (1978) FEBS Lett. 89, 283–286.

[4] Rinderknecht, E. and Humbel, R.E. (1978) J. Biol. Chem. 253, 2769–2775.

[5] Marquardt, H., Todaro, G.J., Henderson, L.E. and Oroszlan, S. (1981) J. Biol. Chem. 256, 6859–6865.

[6] Bell, G.I., Merryweather, J.P., Sanchez-Pescado, R., Stempien, M.M., Priestly, L., Scott, J. and Rall, L.B. (1984) Nature 310, 775–777.

[7] Brissenden, J.E., Ullrich, A. and Francke, U. (1984) Nature 310, 781–784.

[8] Tricoli, J.V., Rall, L.B., Scott, J., Bell, G.I. and Shows, T.B. (1984) Nature 310, 784–786.

[9] Höppener, J.W.M., De Pagter-Holthuizen, P., Geurts van Kessel, A.H.M., Jansen, M., Kittur, S.D., Antonarakis, S.E., Lips, C.J.M. and Sussenbach, J.S. (1985) Hum. Genet. 69, 157–160.

[10] De Pagter-Holthuizen, P., Höppener, J.W.M., Jansen, M., Geurts van Kessel, A.H.M., Van Ommen, G.J.B. and Sussenbach, J.S. (1985) Hum. Genet. 69, 170–173.

[11] Van Ommen, G.J.B., Arnberg, A.C., Baas, F., Brocas, H., Sterk, A., Tegelaars, W.H.H., Vassart, G. and De Vijlder, J.J.M. (1983) Nucleic Acids Res. 11, 2273–2285.

[12] Bates, B.F. and Swift, R.A. (1983) Gene 26, 137–146.

[13] Benton, W.D. and Davis, R.W. (1977) Science 196, 180–182.

[14] Birnboim, H.C. and Doly, J. (1979) Nucleic Acids Res. 5, 1513–1523.

[15] Maxam, A.M. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560–564.

[16] Ullrich, A., Berman, C.H., Dull, T.J., Gray, A. and Lee, J.M. (1984) EMBO J. 3, 361–364.

[17] Dull, T.J., Gray, A., Hayflick, J.S. and Ullrich, A. (1984) Nature 310, 777–781.

[18] Breathnach, R. and Chambon, P. (1981) Annu. Rev. Biochem. 50, 349–383.

[19] Atweh, G.T., Anagnou, N.P., Shearin, J., Forget, B.G. and Kaufman, R.E. (1985) Nucleic Acids Res. 13, 777–790.

[20] 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.