Nucleotide sequences of cDNAs encoding precursors of human insulin-like growth factor II (IGF-II) and an IGF-II variant

M. Jansen, F.M.A. van Schaik*+, H. van Tol*, J.L. Van den Brande and J.S. Sussenbach*

Department of Pediatrics and **Laboratory for Physiological Chemistry, State University of Utrecht, Vondellaan 24[a], 3521 GG Utrecht, The Netherlands*

Received 29 October 1984

We have isolated 3 cDNA clones encoding human IGF-II and a variant of IGF-II. The amino acid sequence encoded by the IGF-II cDNA is identical to the sequence previously described [(1978) FEBS Lett. 89, 283–286]. In the amino acid sequence predicted by the IGF-II variant cDNA, the Ser residue 29 in the B-domain has been replaced by an Arg-Leu-Pro-Gly sequence. The corresponding mRNAs probably arise by alternative splicing of a common RNA precursor. The IGF coding region of the cDNA inserts is flanked by sequences encoding a signal peptide and a carboxy-terminal peptide indicating that both human IGF-II and its variant are synthesized as precursors.

Somatomedin *Insulin-like growth factor II* *Nucleotide sequence* *Amino acid sequence* *Hormone precursor* *cDNA cloning*

1. INTRODUCTION

The somatomedins (SM) or insulin-like growth factors (IGF) constitute a heterogeneous family of peptides with insulin-like and growth-promoting effects [1]. The amino acid sequences of the two main types of IGF in human serum, IGF-I – which is identical to SM-C [2] – and IGF-II – which is very similar to multiplication-stimulating activity in the rat [3] – show 62% homology [4,5]. However, these two somatomedins may have a quite different physiology. IGF-I levels in blood are much more under growth hormone control than IGF-II and, while both have mitogenic effects in vitro, there is suggestive evidence that IGF-II is more important in antenatal growth [6–8]. We have reported the nucleotide sequence of a human liver cDNA encoding IGF-I/SM-C flanked by amino- and carboxy-terminal extensions, thus providing evidence that IGF-I/SM-C is synthesized as a precursor [9]. We now report the isolation and characterization of 3 cDNAs, isolated from the same human liver cDNA library, encoding IGF-II as well as a variant of IGF-II.

+ To whom correspondence should be addressed

2. MATERIALS AND METHODS

2.1. *cDNA library*

The adult human liver cDNA library, which we also used for the isolation of the IGF-I/SM-C, was kindly provided by Dr D. Woods of Children's Hospital Medical Center, Boston, MA, USA [10].

2.2. *Screening procedure*

As a probe for screening the library we used the *Bam*HI-*Pst*I restriction fragment (540 base pairs) of the cDNA encoding human IGF-I/SM-C [9], which contains the entire coding sequence for this growth factor. This cDNA fragment was labeled with ^{32}P by nick-translation [11].

Colonies were grown on nitrocellulose filters, replicated and screened according to the colony hybridization method in [12] with minor modifications. After overnight hybridization at 42°C in a

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/85/$3.30

mixture containing 40% formamide, 5 × Denhardt's solution (0.1% Ficoll, 0.1% polyvinylpyrrolidone, 0.1% bovine serum albumin), 0.75 M NaCl, 75 mM Na-citrate, 0.1% SDS and 100 μg/ml calf thymus DNA, the nitrocellulose filters were washed at 60°C in SSC (0.15 M NaCl, 15 mM Na-citrate) for 1 h and autoradiographed. Subsequently, they were washed in 0.3 × SSC at 60°C for ½ h and again autoradiographed.

2.3. *Analysis of cDNA*

Plasmid cDNA was isolated according to the alkaline lysis method in [13]. Nucleotide sequence analysis was performed as in [14].

3. RESULTS

The amino acid sequences of IGF-I/SM-C and IGF-II show 62% homology, hence the IGF-I/SM-C cDNA could be expected to cross-hybridize with IGF-II cDNA under conditions of reduced stringency. Four colonies were identified, showing hybridization to the ^{32}P-labeled IGF-I/SM-C cDNA probe under such conditions. Only one of them still hybridized with the probe after the wash at increased stringency (0.3 × SSC, 60°C, ½ h). Restriction endonuclease analysis of the plasmid isolated from this clone showed a pattern similar to that of the plasmid containing IGF-I cDNA [9]. The other 3 cDNA inserts were very similar to each other in their restriction pattern, but differed from that of the IGF-I/SM-C cDNA. These inserts (A, B, C) were subjected to nucleotide sequence analysis following the strategy depicted in fig.1. The nucleotide sequences of the cDNA strands homologous to the corresponding mRNAs are shown in fig.2. The 660 nucleotide long cDNA insert of clone A encodes the first 160 amino acid residues of human preproinsulin-like growth factor II [15].

The nucleotide sequences of the cDNA inserts of clones B and C partially overlap and complement each other. The combined sequence comprises 846 nucleotides, containing an open reading frame of 650 nucleotides between the TGA termination codons at position −103 and 550, respectively. The predicted amino acid sequence is identical to that of human pre-pro-IGF-II [15], with the exception of the Ser residue at position 29 in the B-domain of IGF-II [4], which has been replaced by the tetrapeptide Arg-Leu-Pro-Gly. We therefore refer to this 70 amino acid long IGF-II related peptide as an IGF-II variant. The nucleotide sequence is otherwise identical to that of clone A, as far as they overlap. These results indicate that the IGF-II variant is synthesized as a precursor with a total length of 183 amino acids, and an M_r of 20477, while the IGF-II precursor is 180 amino acids long and has an M_r of 20141.

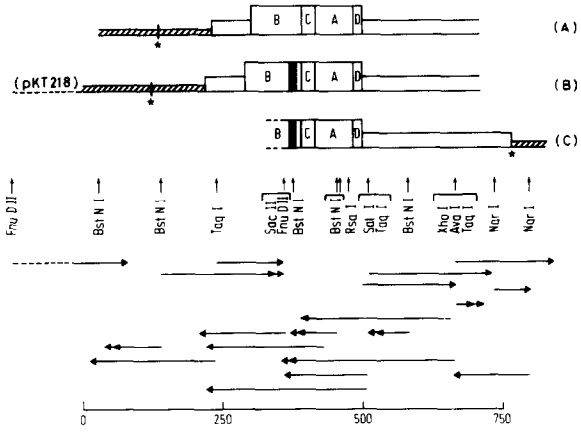

Fig.1. Organization and sequencing strategy for the 3 cDNAs encoding IGF-II (A) and the IGF-II variant (B,C). The regions encoding the B, C, A and D-domains (higher open boxes) as well as those coding for the signal peptide and the C-terminal peptide (lower open boxes) are depicted schematically. The black bar in the B-domain of clones B and C indicates the nonanucleotide insert in the variant-IGF-II sequence. The 5′ and 3′ non-translated regions are hatched; asterisks indicate stopcodons. Of the restriction sites only those used for sequencing are indicated. The horizontal arrows indicate the direction and extent of the sequence determinations; single-headed arrows refer to clones B and C, whereas double-headed arrows indicate confirmatory determinations in clone A.

4. DISCUSSION

From the same human liver cDNA library we have isolated two closely related but distinct types of cDNA encoding precursors of IGF-II and a variant of IGF-II. Both precursors appear to have identical amino-terminal and carboxyterminal peptides. During the preparation of this report, the complete nucleotide sequence of pre-pro-IGF-II was published [15] for another cDNA isolated

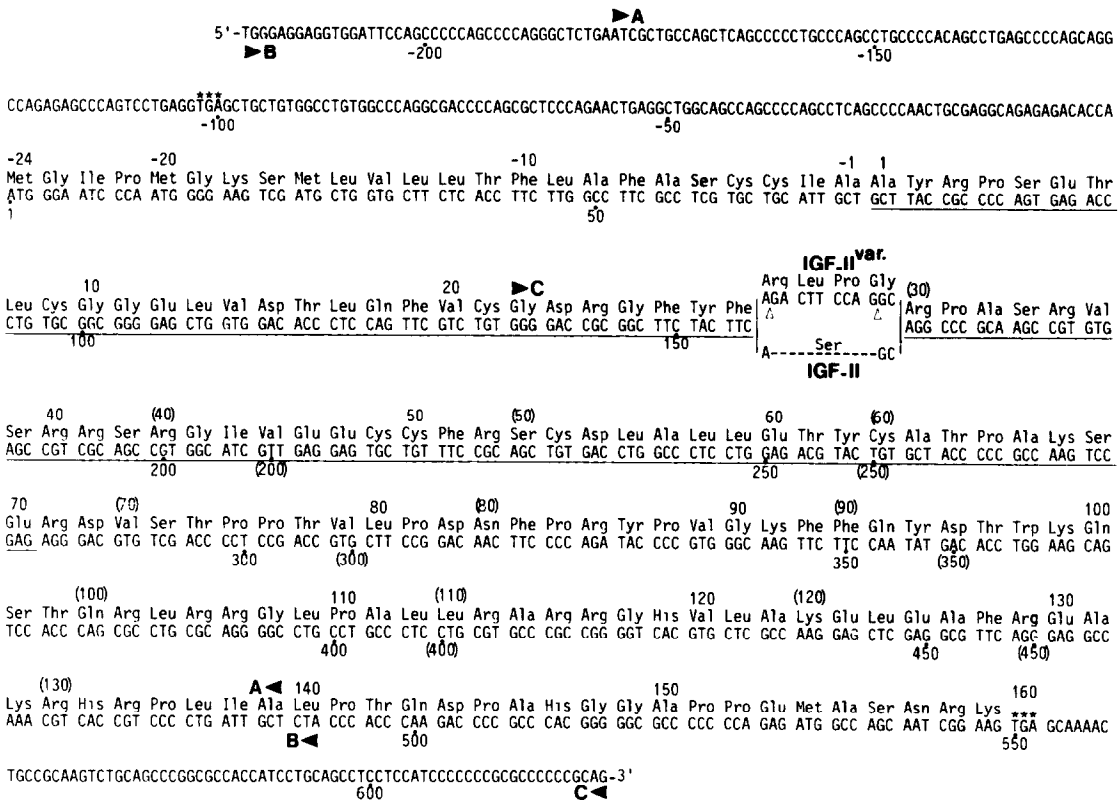

Fig.2. Nucleotide sequence of the cDNA-inserts encoding human IGF-II and IGF-II variant. The DNA strand presented is homologous to the IGF-mRNAs. Nucleotides are numbered sequentially from the ATG codon encoding the Met residue at position −24; nucleotides preceding this ATG codon are given negative numbers. The predicted amino acid sequence is numbered starting at the amino-terminus of the IGF-sequences (underlined); amino acid residues of the signal peptide are given negative numbers. Following the region in the B-domain where the sequences of IGF-II and its variant diverge, the amino acid and nucleotide numbers referring to IGF-II are bracketed. The black arrowheads plus letters indicate the determined sequences of the cDNA-inserts from clones A, B and C. The sequence between the open arrowheads represents the nonanucleotide insert in the IGF-II variant. Asterisks indicate the termination codons, TGA.

from the same library. Both nucleotide sequences are identical, except for an additional C at position −80, which brings the termination codon TGA at nucleotide position −103 in the same reading frame as the initiation codon ATG as position 1. Comparison of the nucleotide sequences of the regions coding for IGF-II and the IGF-II variant demonstrates that the IGF-II variant cDNA contains a short insert of 9 nucleotides at nucleotide positions 158–166. Theoretically, the presence of two closely related cytoplasmic mRNAs in the original liver cells corresponding to both types of IGF-II can be explained by postulation of two separate genes for IGF-II and the variant. However, circumstantial evidence from chromosomal mapping studies as performed by our group [16] as authors in [17] and [18] points to the existence of only a single IGF-II gene. Another explanation might be that the two species of mRNA arise by alternative splicing of a common RNA precursor. The presence of a typical acceptor site (5′-CTTCCAG-3′) in the sequence of the extra 9 nucleotides of the variant IGF-II cDNA strongly favors the latter explanation. Moreover, nucleotide sequence analysis of the IGF-II gene as reported in [19] indicates the presence of an intervening sequence of approx. 1700 nucleotides exactly at this point, the A and G nucleotides of the

AGC codon (Ser B 29) functioning as donor and acceptor splice site, respectively. The last 9 nucleotides of this intron perfectly match the nonanucleotide insert in the IGF-II variant precursor. Whether this alternative splicing is tissue-specific is unknown and requires further investigation.

ACKNOWLEDGEMENT

We thank Dr D. Woods of Children's Hospital Medical Center, Boston, MA 02115, USA, for providing us with the Adult Human Liver cDNA library.

REFERENCES

[1] Daughaday, W.H., Hall, K., Rabin, M.S., Salmon, W.D., Van den Brande, J.L. and Van Wyk, J.J. (1972) Nature 235, 107.
[2] Klapper, D.G., Svoboda, M.E. and Van Wyk, J.J. (1974) Endocrinology, 112, 2215–2217.
[3] Marquardt, H., Todaro, G.J., Henderson, L.E. and Oroszlan, S. (1981) J. Biol. Chem. 256, 6859–6865.
[4] Rinderknecht, E. and Humbel, R.E. (1978) FEBS Lett. 89, 283–286.
[5] Rinderknecht, E. and Humbel, R.E. (1978) J. Biol. Chem. 253, 2769–2775.
[6] Daughaday, W.H. (1972) Proc. Soc. Exp. Biol. Med. 170, 257–263.
[7] Zapf, J., Schmid, C. and Froesch, E.F. (1984) Clin. Endocrinol. Metab. 13, 3–30.
[8] Underwood, L.E. and D'Ercole, A.J. (1984) Clin. Endocrinol. Metab. 13, 69–89.
[9] 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.
[10] Woods, D.E., Markham, A.F., Ricker, A.T., Goldberger, G. and Colten, H.R. (1982) Proc. Natl. Acad. Sci. USA 79, 5661–5665.
[11] Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg, P. (1977) J. Mol. Biol. 113, 237–251.
[12] Grunstein, M. and Hogness, D.S. (1975) Proc. Natl. Acad. Sci. USA 72, 3961–3965.
[13] Birnboim, H.C. and Doly, J. (1979) Nucleic Acids Res. 5, 1513–1523.
[14] Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499–560.
[15] Bell, G.I., Merryweather, J.P., Sanchez-Pescador, R., Stempien, M.M., Priestley, L., Scott, J. and Rall, L.B. (1984) Nature 310, 775–777.
[16] 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. (1984) submitted.
[17] Brissenden, J.E., Ullrich, A. and Francke, U. (1984) Nature 310, 781–784.
[18] Tricoli, J.V., Rall, L.B., Scott, J., Bell, G.I. and Shows, T.B. (1984) Nature 310, 784–786.
[19] Dull, T.J., Gray, A., Hayflick, J.S. and Ullrich, A. (1984) Nature 310, 777–781.