Volume 214, number 2, 259–264 FEB 04595 April 1987

The human insulin-like growth factor II gene contains two development-specific promoters

P. de Pagter-Holthuizen, M. Jansen+, F.M.A. van Schaik, R. van der Kammen+, C. Oosterwijk, J.L. Van den Brande+ and J.S. Sussenbach

Laboratory for Physiological Chemistry and +*Department of Pediatrics, State University of Utrecht, Vondellaan 24A, 3521 GG Utrecht, The Netherlands*

Received 23 January 1987

The insulin-like growth factors (IGF) play an important role in fetal and postnatal development. Recently, the nucleotide sequences of the cDNAs encoding IGF-I and IGF-II and part of the human IGF genes were reported. In this communication we describe two distinct IGF-II cDNAs isolated from a human adult liver and a human hepatoma cDNA library, respectively. Using these two cDNAs, we have established that the human IGF-II gene contains at least 7 exons. Two different IGF-II promoters have been identified, 19 kilobases (kb) apart, which are active in a development-specific manner. The promoter, active in the adult stage, is located only 1.4 kb downstream from the insulin gene.

Insulin-like growth factor II; Somatomedin; Eukaryotic promoter; Development; (Human gene)

1. INTRODUCTION

The somatomedins or insulin-like growth factors (IGF) play an important role in fetal and postnatal growth and development [1,2]. Two major human IGFs have been fully characterized; IGF-I is a basic peptide of 70 amino acids, while IGF-II is a neutral peptide containing 67 residues [3,4]. The nucleotide sequences of cDNAs encoding IGF-I and IGF-II and part of the human IGF genes have been reported [5–14]. The cDNA sequences predict that both IGFs are synthesized as larger precursor molecules which undergo extensive processing.

Here, we describe the characterization of two distinct IGF-II cDNAs isolated from a human adult liver and a human hepatoma cDNA library, respectively. These cDNAs, differing from each other in their 5′-noncoding regions, have been used to determine the gene structure. We have identified two different IGF-II promoters, 19 kb apart, which are active in a development-specific manner.

Correspondence address: P. de Pagter-Holthuizen, Laboratory for Physiological Chemistry, State University of Utrecht, Vondellaan 24A, 3521 GG Utrecht, The Netherlands

2. MATERIALS AND METHODS

2.1. *Libraries*

A cDNA library from the human hepatoma cell line HepG2 was kindly provided by Drs P. Berg and M. McPhaul (Stanford, USA). It was constructed as described in [15] and has a complexity of 2×10^6 independent clones. Two non-amplified human genomic cosmid libraries were constructed from human placenta DNA and from GM 1416, 48XXXX cell line DNA [11].

2.2. *Restriction mapping, nucleotide sequence analysis and primer extension*

The isolated cDNAs and cosmid clones were characterized by restriction enzyme analysis and Southern blot hybridization. Using [α-^{32}P]dCTP-

Published by Elsevier Science Publishers B.V. (Biomedical Division)

labeled cDNAs as probes, restriction maps were constructed. For nucleotide sequence analysis the procedure of Maxam and Gilbert [16] was used. Primer extension was performed as in [17].

2.3. *DNA probes*

The cDNA probes isolated from the plasmid pIGF-II [8] contain the coding sequences of the IGF-II precursor, preceded by a 5′-nontranslated region of 526 bp. Genomic probes were isolated from the cosmid clones and contain sequences specific for several exons as indicated in the legend to fig.3. Double-stranded DNA probes were labeled by nick translation with [α-^{32}P]dCTP to a specific activity of 10^8 cpm/μg.

2.4. *Northern blotting*

Poly(A)$^+$ RNA was isolated from fetal and adult liver by the guanidinium thiocyanate/CsCl method [18] and one round of oligo(dT)-cellulose chromatography [19]. Glyoxalated RNAs [20] were size-fractionated on 0.8% agarose gels (10 μg per lane), transferred onto nylon hybridization membranes (Hybond N, Amersham, England) and hybridized to different ^{32}P-labeled probes. After hybridization the membranes were washed to a final stringency of 0.1 × SSC, 1% SDS, at 65°C for 30 min.

3. RESULTS AND DISCUSSION

From an adult human liver cDNA library several IGF-II-specific cDNAs have been isolated, but none contains the 5′-terminus of IGF-II mRNA [8]. Localization of the cDNA sequences on cosmid clones showed that the IGF-II gene contains at least two 5′-noncoding exons and three coding exons [11]. Further analysis revealed that one of these cDNAs contains sequences derived from a third 5′-noncoding exon. A schematic representation of this cDNA and the nucleotide sequence of its 5′-terminus are shown in fig.1A. These results imply that IGF-II mRNA from adult liver is transcribed from six exons, viz. three 5′-nontranslated exons (exons 1–3) and three exons coding for the IGF-II precursor (exons 5–7).

From the human hepatoma cell line HepG2 we isolated one IGF-II-specific clone. Analysis of this cDNA revealed that it consists of the three coding exons (exons 5–7) preceded at the 5′-end by a

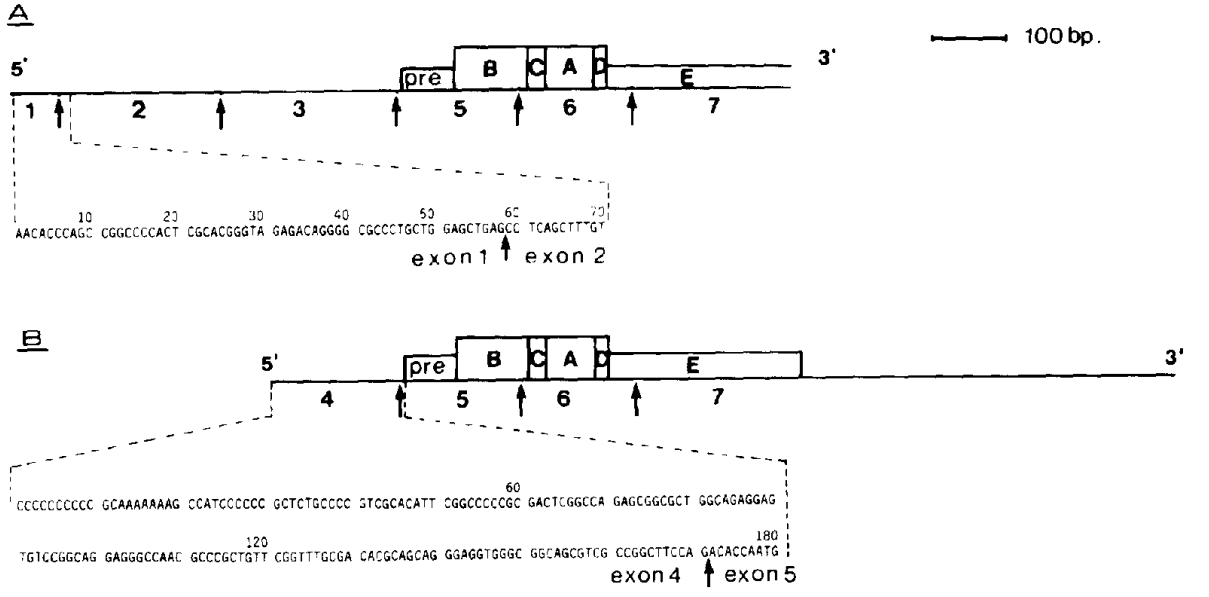

Fig.1. Schematic representation and 5′-nucleotide sequences of IGF-II cDNAs isolated from a human adult liver library (A) and a human hepatoma cell line HepG2 library (B). The nucleotide sequences of exons 2–7 have been published elsewhere [7,8,11]. The regions encoding the mature peptide (B,C,A,D), those coding for the signal peptide (pre) and the C-terminal peptide (E) as well as 5′- and 3′-nontranslated sequences are indicated.

nucleotide sequence, diverging from the splice site onwards, which has not been detected before in other human cDNAs (fig.1B). The sequence is homologous to a rat liver cell line (BRL3A) cDNA sequence determined by Dull et al. [7] and identical to a sequence localized on the human chromosome. This 5′-noncoding sequence is designated exon 4.

Employing the two above-mentioned IGF-II cDNAs as probes we determined the precise positions of the IGF-II specific sequences on chromosomal cosmid clones. The complete map of the IGF-II gene is shown in fig.2. The human IGF-II gene contains seven exons and spans 28 kb of chromosomal DNA. Four 5′-noncoding exons (exons 1–4) are followed by three exons (exons 5–7) coding for the IGF-II precursor. The 5′-noncoding exons are used in alternative transcripts. Adult liver mRNA consists of exons 1–3, 5–7, while HepG2 mRNA contains exons 4–7, suggesting the presence of two different promoters. A striking feature of the IGF-II gene is its location close to the insulin gene. Bell and co-workers [10] established that the genes for insulin and IGF-II are contiguous with a maximal distance of 12.6 kb. Our data further reduce this distance to only 1.4 kb (fig.2).

Since the two characterized cDNAs contain different 5′-nontranslated sequences, while only a single IGF-II gene is present, we established at which stage in development the two corresponding mRNAs are expressed. Poly(A)$^+$ RNA was isolated from fetal and adult human liver. Northern blots of poly(A)$^+$ RNA were hybridized with three different ^{32}P-labeled probes. Hybridization with an IGF-II probe containing exon 6 to RNA blots containing 10 μg poly(A)$^+$ RNA from fetal and adult human liver revealed strong expression of a 6.0 kb mRNA in fetal liver, while in adult liver mRNA a weak band of 5.3 kb was detected (fig.3, lanes 3,4). This indicates that the IGF-II gene is predominantly expressed in fetal tissue. To establish the expression of the different 5′-noncoding exons, poly(A)$^+$ RNA blots were hybridized with fragments containing exon 1 and exon 4 sequences, respectively. The probe containing exon 1 sequences hybridized to the 5.3 kb band in adult mRNA and not to the 6.0 kb band in fetal liver (fig.3, lanes 5,6) indicating that exon 1 sequences are only present in adult mRNA and not in fetal mRNA. On the other hand, the exon 4 probe hybridized only to the 6.0 kb band in fetal mRNA (fig.3, lanes 1,2), suggesting that this exon is only expressed in fetal tissue.

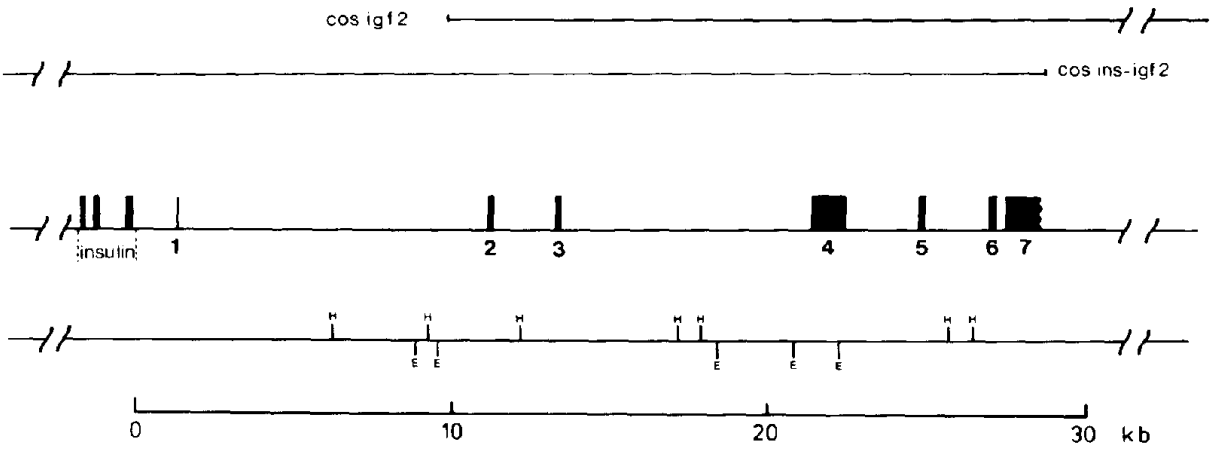

Fig.2. Schematic representation of the insulin and IGF-II gene region. Two partially overlapping cosmid clones were used to characterize the IGF-II gene structure: *cos igf2* contains exons 2–7 of the IGF-II gene [11] and *cos ins-igf2*, isolated from the GM1416 cosmid library, encompasses the seven exons of the IGF-II gene as well as the three exons of the insulin gene. The positions of the IGF-II exons were determined by restriction enzyme and nucleotide sequence analysis. The structure and sequence of the insulin gene have been reported previously by Bell et al. [26]. A restriction map of the IGF-II and insulin gene region for the restriction enzymes *Hin*dIII (H) and *Eco*RI (E) is shown.

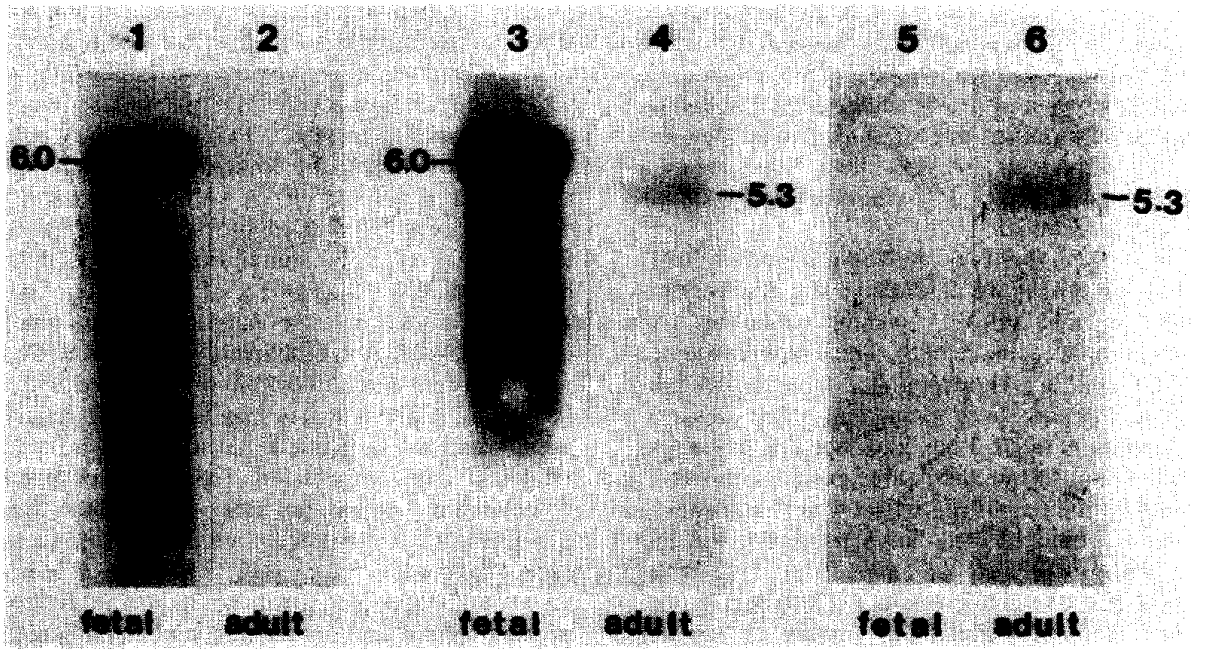

Fig.3. Northern blots of fetal and adult human liver RNA. Poly(A)$^+$ RNA was isolated from human fetal and adult liver (10 μg per lane). Lanes: 1,2, hybridization with an 852 bp genomic fragment of exon 4 (nucleotides 164–915 in [7]); 3,4, hybridization with an 825 bp genomic fragment containing sequences of exon 6, coding for mature peptide and the initial sequences of the E-domain; 5,6, hybridization with a 945 bp genomic fragment (nucleotides 503–1447, fig.4B) containing part of the exon 1 sequence as well as upstream sequences.

To determine the precise positions of initiation of transcription, primer extension experiments were performed with poly(A)$^+$ RNA from fetal and adult liver. A primer of 20 nucleotides (complementary to nucleotides 272–291, fig.4A) hybridized to fetal mRNA, resulting in an elongation product of about 100 nucleotides (not shown). Nucleotide sequence analysis of the primer extension product revealed that initiation of transcription in the fetal stage occurs at position 186 (fig.4A). An analogous experiment with adult poly(A)$^+$ RNA and an exon 1 primer (complementary to nucleotides 1469–1488, fig.4B) yielded extension products of 90–95 nucleotides, suggesting that initiation of transcription takes place at nucleotides 1394–1397 (fig.4B). Due to the low abundance of IGF-II mRNA in adult liver, nucleotide sequence analysis of the primer extension product was not feasible.

The identification of two sites of initiation of transcription suggests the presence of two independent promoters. In order to investigate the structure of these promoters we have determined the genomic nucleotide sequences preceding the two start sites of transcription. For the fetal promoter, the region upstream of exon 4 was sequenced (1–185, fig.4A). In this region a TATA box is found at positions -25 to -19 with respect to the site of initiation of transcription, while a CAAT box (-82 to -86) and an Sp1 recognition sequence (-113 to -104) are also present. These sequences are characteristic elements of most eukaryotic promoters [21].

For characterization of the adult promoter we have determined the nucleotide sequence of the 1.4 kb intergenic region between the insulin gene and exon 1 of the IGF-II gene (fig.4B). This region exhibits a number of remarkable features. Firstly, the region upstream of exon 1 does not contain TATA and CAAT boxes, but an Sp1 recognition site (nucleotides 1344–1353) is present. Further, a GC-rich region of about 80 nucleotides precedes the putative site of initiation of transcription. These features have also been established for a

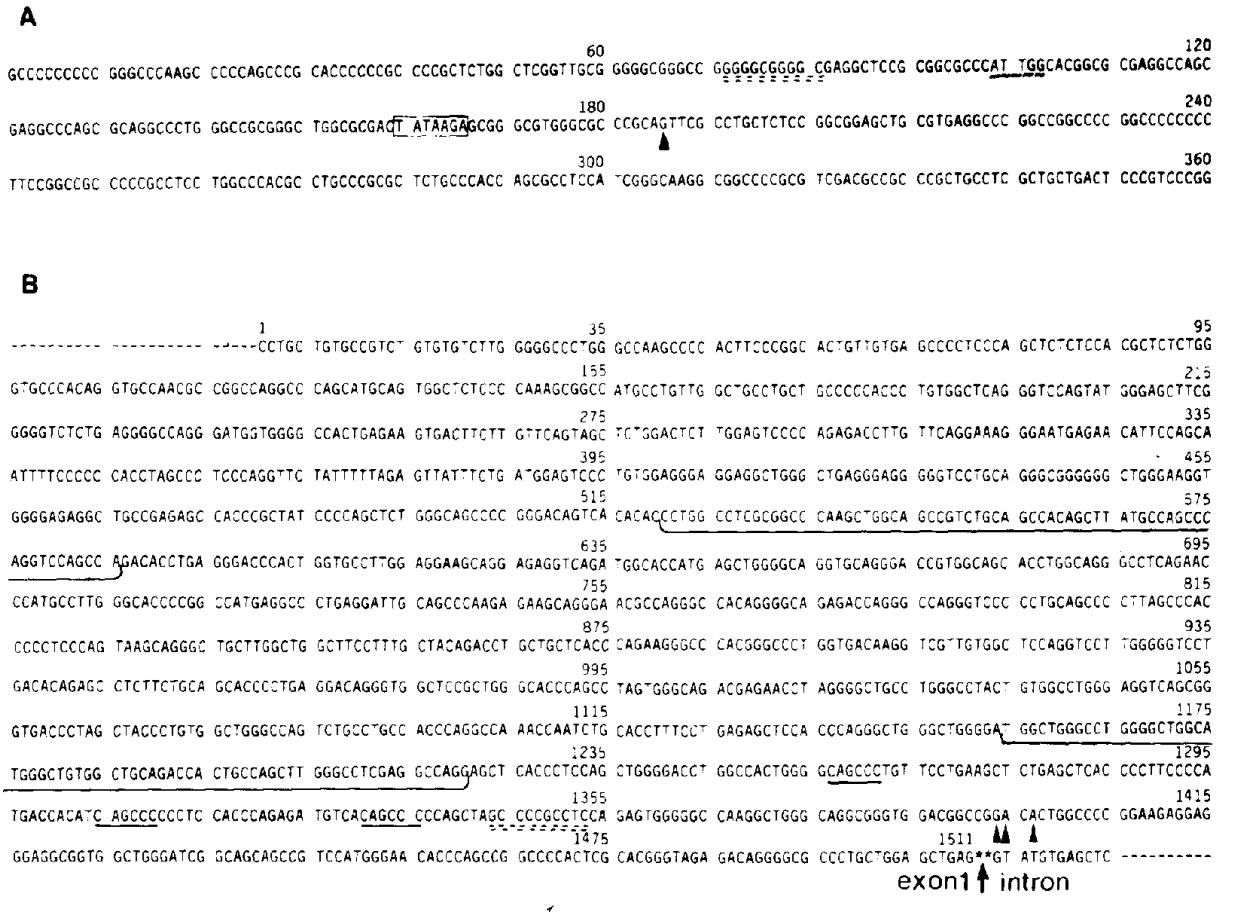

Fig.4. Nucleotide sequences of human chromosomal DNA containing the fetal and adult IGF-II promoters. (A) Genomic nucleotide sequence containing part of exon 4 and its 5′-flanking region. Nucleotides 258–360 were also determined by Dull et al. [7]. This region contains the IGF-II promoter active in human fetal liver. The TATA sequence is boxed, while a CAAT-box (~~~) and an Sp1 recognition sequence (====) are indicated. The site of transcription initiation is shown (▲). (B) Nucleotide sequence of the intergenic region between the human insulin and IGF-II genes containing the IGF-II promoter active in human adult liver. Nucleotide 1 is the first nucleotide downstream of the polyadenylation site of the insulin gene [25]. The site of initiation of transcription determined by primer extension is indicated (▲). The repeated CAGCCC sequence is underlined as well as a 66 bp inverted repetition. An Sp1 recognition sequence is shown (====).

number of so-called housekeeping genes, which are expressed at low levels in a variety of tissues [22]. Secondly, the intergenic region contains a number of direct and inverted repeats. Besides several CAGCCC repeats there is an almost perfect 66 bp inverted repeat at positions 521–586 and 1155–1221 (fig.4B). These repeats might be involved in regulation of expression by interaction with regulatory proteins.

Since the nucleotide sequences upstream of exon 1 do not contain typical eukaryotic promoter elements we have investigated whether this region exhibits promoter activity. A 267 bp fragment (nucleotides 1186–1452) of the intergenic region was inserted in front of a promoter-defective neomycin-resistance gene. A mouse hepatoma cell line [23] was transfected with this eukaryotic expression plasmid and the transfected cells were tested for expression of the neomycin-resistance gene. Preliminary results show that the 267 bp fragment acts as a strong promoter (not shown). This confirms the notion that the region upstream

from exon 1 is involved in expression of the IGF-II gene in adult liver. Further analysis of the promoter regions of the IGF-II gene is in progress.

Tissue-specific expression of the IGF-II gene has been reported previously [24]. The present data indicate that the IGF-II gene can also be expressed in a development-dependent way using two different promoters. Since the number of known genes with development-specific expression is still very limited [25], more detailed investigation of the human IGF-II gene with its remarkable regulation of expression is imperative. This might lead to the identification of development-specific transcription factors and the elucidation of various processes involved in development and differentiation.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge Drs P. Berg and M. McPhaul (Stanford, USA) for the HepG2 cDNA library, Dr G.J.B. van Ommen (Leiden, The Netherlands) for cosmid libraries and Dr G.I. Bell (Emeryville, USA) for communicating unpublished sequence data. This research was supported in part by the Foundation for Medical Research (FUNGO) with financial aid from the Netherlands Organization for the Advancement of Pure Research (ZWO).

REFERENCES

[1] Daughaday, W.H. (1984) Clinics in Endocrinology and Metabolism, Tissue Growth Factors, vol.13, pp.3–189, W.B. Saunders, Philadelphia.
[2] Copinschi, G. and Chatelain, P. (1986) Horm. Res. 24, 81–200.
[3] Rinderknecht, E. and Humbel, R.E. (1978) J. Biol. Chem. 253, 2769–2776.
[4] Rinderknecht, E. and Humbel, R.E. (1978) FEBS Lett. 89, 283–286.
[5] 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.
[6] 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.
[7] Dull, T.J., Gray, A., Hayflick, J.S. and Ullrich, A. (1984) Nature 310, 777–781.
[8] 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.
[9] 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.
[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] LeBouc, Y., Dreyer, D., Jaeger, F., Binoux, M. and Sondermeyer, P. (1986) FEBS Lett. 196, 108–112.
[13] Rotwein, P. (1986) Proc. Natl. Acad. Sci. USA 83, 77–81.
[14] Rotwein, P., Pollock, K.M., Didier, D.K. and Krivi, G.G. (1986) J. Biol. Chem. 261, 4828–4832.
[15] Okayama, H. and Berg, P. (1983) Mol. Cell. Biol. 3, 280–289.
[16] Maxam, A. and Gilbert, W. (1980) Methods Enzymol. 65, 499–560.
[17] Toole, J.J., Knopf, J.L., Wozney, J.M., Sultzman, L.A., Buecker, J.L., Pitman, D.D., Kaufman, R.J., Brown, E., Shoemaker, C., Orr, E.C., Amphlett, G.W., Foster, W.B., Coe, M.L., Knutson, G.J., Fass, D.N. and Hewick, R.M. (1984) Nature 312, 342–347.
[18] Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) Biochemistry 18, 5294–5299.
[19] Aviv, H. and Leder, P. (1972) Proc. Natl. Acad. Sci. USA 69, 1408–1412.
[20] Thomas, P.S. (1983) Methods Enzymol. 100, 255–266.
[21] McKnight, S. and Tjian, R. (1986) Cell 46, 795–805.
[22] Dynan, W.S. (1986) Trends Genet. 2, 196–197.
[23] Darlington, G.J., Bernhard, H.P., Miller, R.A. and Ruddle, F. (1980) J. Natl. Cancer Inst. 64, 809–819.
[24] 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.
[25] Leff, S.E., Rosenfeld, M.G. and Evans, R.M. (1986) Annu. Rev. Biochem. 55, 1091–1117.
[26] Bell, G.I., Pictet, R.L., Rutter, W.J., Cordell, B., Tischer, E. and Goodman, H.M. (1980) Nature 284, 26–32.