Volume 13 Number 4 1985 Nucleic Acids Research

Developmental and tissue-specific expression of a family of transcripts related to rat insulin-like growth factor II mRNA

Marcelo Bento Soares, Douglas N. Ishii* and Argiris Efstratiadis

Departments of Human Genetics and Development and *Pharmacology, Columbia University, 701 West 168th Street, New York, NY 10032, USA

Received 14 December 1984; Accepted 24 January 1985

ABSTRACT

We have constructed a cDNA library from the mRNA of a rat liver cell line (BRL-3A) and characterized cDNA clones encoding the protein precursor of the rat insulin-like growth factor II (pre-pro-rIGF-II). This precursor, inferred from the nucleotide sequence, consists of a signal peptide, the rIGF-II sequence, and a trailer polypeptide of unknown significance. The characterized cDNA sequence (1016 nt) is part of a 3.4 kb mRNA species. Northern analysis reveals that a probe containing the extreme 5' noncoding region hybridizes to a second RNA (1.6 kb), while a probe corresponding to the 5' noncoding region proximal to the coding region hybridizes to two other RNA species (1.75 and 1.1 kb). All four RNAs are differentially expressed in all of the neonatal tissues that were examined, while the 3.4 kb pre-pro-rIGF-II mRNA and the 1.1 kb transcript are absent from adult tissues.

INTRODUCTION

The somatomedins (SMs) or insulin-like growth factors (IGFs), which have a structural resemblance to proinsulin, are mitogenic polypeptides that were thought to mediate the effects of growth hormone (somatotropin) on skeletal tissue (see ref.1-4 for reviews). Though this may be true for one of the two known SMs (IGF-I), the other (IGF-II), which is presumably involved in fetal development, seems to be under the influence of placental lactogen, but not of growth hormone (5,6).

The structures of human IGF-I (hIGF-I or SM-C) and IGF-II (hIGF-II or SM-A) have been determined (7-9). A rat basic SM, isolated from serum, has been partially sequenced (10), and it seems that it corresponds to hIGF-I. A second rat SM (previously designated "multiplication-stimulating activity", MSA), purified from serum-free medium conditioned by the Buffalo rat liver cell line BRL-3A (11), was shown by sequencing (12) to be the rat equivalent of hIGF-II (rIGF-II).

Peptide alignments with proinsulin (B chain-C peptide-A chain), show that the following regions can be defined in the sequences of the IGFs: Domains B, C and A. The IGFs also contain a fourth domain, D, which is not present in

proinsulin (13). The limited amino acid sequence homology between the two types of molecules is restricted to the B and A domains. In contrast to insulin, which is generated from proinsulin by elimination of the middle C peptide, the C domain of the SMs remains in the active form of the molecules.

The structural similarities between proinsulin and IGFs suggested common ancestry of the corresponding genes (1). Thus, to expand our previous studies on the evolution (14-16) and expression (17) of insulin genes to other putative members of the insulin-like gene family, we cloned rIGF-II cDNA. Here we describe the characterization of the mRNA species which encodes the protein precursor of rIGF-II and present preliminary observations on other transcripts that partially overlap the sequence of this message. We also show that the family of these transcripts exhibits developmental and tissue specificity.

After completion of this work, two reports appeared describing the characterization of cDNA clones of the same sequence (18,19). However, there are discrepancies between these reports and our data (see below). Regardless of these differences, the molecular characterization of the nucleotide sequences encoding IGFs is well under way: cDNAs and chromosomal genes corresponding to hIGF-I and hIGF-II have been partially characterized (18,20-22).

MATERIALS AND METHODS

Materials

Restriction enzymes were from New England Biolabs or Bethesda Research Laboratories. T4 DNA ligase and Eco RI linkers were from New England Biolabs. S1 nuclease and T4 polynucleotide kinase were from Bethesda Research Laboratories. T4 DNA polymerase was from P-L Biochemicals. Klenow fragment of E.coli DNA polymerase was from Boehringer-Mannheim. Reverse transcriptase was from Life Sciences. Placental ribonuclease inhibitor (RNasin) was from Promega Biotec. Bacterial alkaline phosphatase was from Worthington. Nylon membranes (Gene Screen Plus), α-^{32}P-dNTPs (800 Ci/mmole) and γ-^{32}P-ATP (3000 Ci/mmole) were from New England Nuclear. DNA oligonucleotides were chemically synthesized using the Applied Biosystems 380A DNA Synthesizer. Bacteriophage vector λgt10 and plasmid vector pUC9 were provided by T. Huynh and J. Messing, respectively.

RNA Purification

Cytoplasmic RNA was extracted as described (23) and further purified by oligo(dT)-cellulose chromatography. Total RNA was prepared by the guanidine thiocyanate-CsCl procedure (24).

Recombinant DNA Procedures

Double-stranded cDNA was synthesized as described (23) using as a primer oligo(dT) and as a template cytoplasmic poly(A)$^+$ RNA from BRL-3A cells. The ends of the double-stranded cDNAs were made flush by treatment with T4 DNA polymerase, and the molecules were cloned as described (23) into the vector λgt10, following attachment of Eco RI linkers. Screening of the library was performed as described (25).

The radioactive probe to isolate rIGF-II cDNA was made as follows: We first synthesized a message-like DNA 39mer, corresponding to the first 13 amino acid residues of the A domain of rIGF-II (see Fig.2b). For the third position of the two-fold degenerate codons in the middle of the oligonucleotide we used both of the alternative bases, while for the rest of the codons the base of the third position was selected either at random or by giving advantage to possible G-U base-pairing. We then synthesized a 10mer, complementary to the 3' terminal region of the 39mer. Following annealing of the two oligonucleotides, the 10mer was extended with labeled triphosphates using the Klenow enzyme. It is our experience that this strategy is superior to the use of a mixture of shorter oligonucleotides containing all possible combinations (which, in our hands, has resulted several times in the isolation of fortuitous hybridizers). The success of this approach is due to the higher specific activity of the probe that can be obtained (in contrast to end-labeled oligonucleotides). Moreover, nucleation events in the middle of the probe (where an exact complement in relatively high representation is encountered) minimize the disadvantages of possible mismatches at the two ends.

A chromosomal DNA clone was isolated by screening with cloned cDNAs $6x10^5$ plaques of a genomic rat DNA library (provided by T.Sargent). This library had been constructed by partial digestion of rat chromosomal DNA with Hae III, and attachment of Eco RI linkers to the resulting fragments prior to ligation to λ Charon 4A arms.

RNA and DNA Analyses

For Northern analysis, RNA was electrophoresed on formaldehyde-agarose gels and then transferred onto nylon membranes. Prehybridization and hybridization were as described (26), except when oligonucleotide probes were used (hybridization at 37^{0}C, in 25% formamide). As probes we used labeled oligonucleotides (see above), nick-translated cloned restriction fragments or uniformly labeled single-strands synthesized on M13 templates by a modification of a published procedure (27). The inserts of the isolated cDNA clones were subcloned into M13mp9, M13mp10, or pUC9 for DNA sequencing, which was performed

by the enzymatic method (28), using reverse transcriptase, or, in certain cases, by the chemical method (29), for verification.

RESULTS AND DISCUSSION

Isolation and Characterization of rIGF-II cDNA Clones

For the isolation of rIGF-II cDNA clones we first examined whether a radioactive antimessage probe, synthesized from a synthetic DNA template corresponding to the first 13 amino acid residues of the A domain of rIGF-II (see Materials and Methods and Fig.2b), could hybridize to cytoplasmic poly(A)$^+$ RNA from the BRL-3A cell line. Fig.1 (lane A) shows that an RNA species approximately 3.4 kb in length gave a strong hybridization signal.

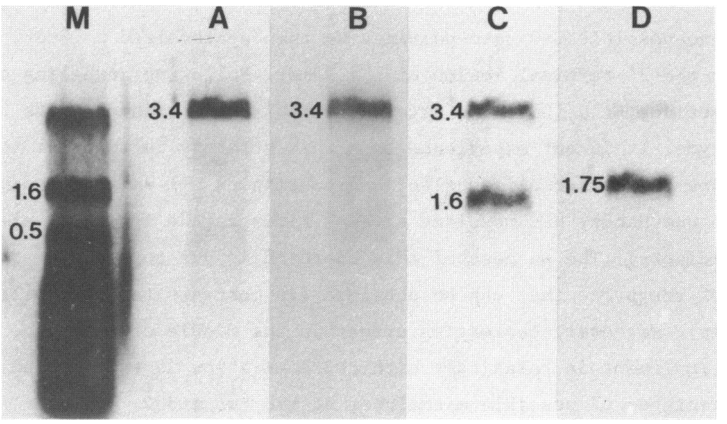

Fig. 1. Northern analysis of cytoplasmic poly(A)$^+$ RNA from BRL-3A cells. The RNA (10 µg per lane) was electrophoresed on a 1% agarose-formaldehyde gel and then transferred onto a nylon membrane. Lanes A and B represent two slots of the same gel. Lanes C and D represent sequential hybridizations with different probes of the same membrane as in A, after decay of the radioactive signal detected in each previous step. The following probes were used: Lane A, a labeled antimessage oligonucleotide (corresponding to the beginning of the A domain), synthesized on a synthetic DNA template (see Fig.2a and b). Lane B, a uniformly labeled single-stranded probe synthesized from a clone 11c (Fig. 2a) template, in M13. Because of the low concentration of the α-^{32}P-dNTPs during this reaction (which was not followed by chase), the synthesis was limited and the probe represented primarily 3' noncoding region (see Fig.2a). Lane C, a probe generated by nick-translation of the Eco RI-Eco RII fragment of clone 27 (see Fig.2a). Lane D, a uniformly labeled single-stranded probe representing an Eco RI-Mbo II fragment of clone 19f (from the beginning of the sequence in Fig.4 to position 274). The faint hybridization of the 3.4 kb mRNA in this lane might be due to contamination of the probe with sequences of clone 19f homologous to this message (see Fig.4). Lane M (size markers): labeled Hinf fragments of pBR322. The slowest migrating band is a partial digestion product.

We then constructed a cDNA library of $1.5x10^6$ recombinants from the BRL-3A RNA into the vector λgt10. Following library amplification, we screened $4.5x10^5$ plaques with the same probe and identified 32 positive plaques, 25 of which were also positive upon rescreening. Sequencing of the insert of one of these clones (clone 14, Fig.2a) identified a region of homology to the synthetic probe and a reading frame corresponding to the known amino acid sequence (12) of the A and D domains of rIGF-II.

Cross-hybridization of a clone 14 probe to the other 24 positive clones (not shown) indicated that 11 of them carried rIGF-II sequence, while the rest were fortuitous hybridizers. Selected overlapping clones of the positive group (Fig.2a) were also sequenced. Northern analysis of BRL-3A RNA using a 3' noncoding region probe from clone 11c showed that the same 3.4 kb RNA species hybridizing to the synthetic probe also hybridized to rIGF-II cDNA sequence (Fig.1, lane B).

A composite of our sequencing data is presented in Fig.2b. Since the total length of the DNA sequence that we have established is 1016 nt, we are missing approximately 2.4 kb of the sequence of the 3.4 kb mRNA. A primer extension experiment (Fig.3) suggested that the 5' end of the sequence is probably close to the 5' end of the 3.4 kb mRNA. As primer for this extension we used a 536 nt long single-stranded antimessage DNA fragment (Fig.2a), with a 3' end mapping at a Pvu II site (position +222 of the sequence). The longest abundant transcript we detected was approximately 1kb in length (Fig.3). Assuming that this product does not correspond to a strong stop for reverse transcriptase, we calculate that the extended sequence (1000-536=464 nt) is only slightly longer than the distance between the Pvu II site and the 5' end of the sequence (444 nt). We note, however, that at least two other very faint bands of extension products, that can be seen only on the original autoradiogram, were longer than 1 kb. Their appearance might be due to rIGF-II-related sequences (see below). Thus, we tentatively conclude that most of the unknown sequence of the 3.4 kb mRNA corresponds to 3' noncoding region.

By identifying the rIGF-II residues in the sequence of Fig.2b, and extending the translation frame upstream, we assigned the Met at amino acid position -24 as the initiator. We believe that this interpretation is correct because no other ATG triplets are present in the sequence between the -24 Met and the first upstream terminator (TGA at position -144) in the same reading frame. Moreover, the first 23 amino acid residues of this precursor, which presumably serve as a signal peptide for the secretory rIGF-II, contain a hydrophobic core, characteristic of all known preregions (30). Thus, the putative pre-pro-

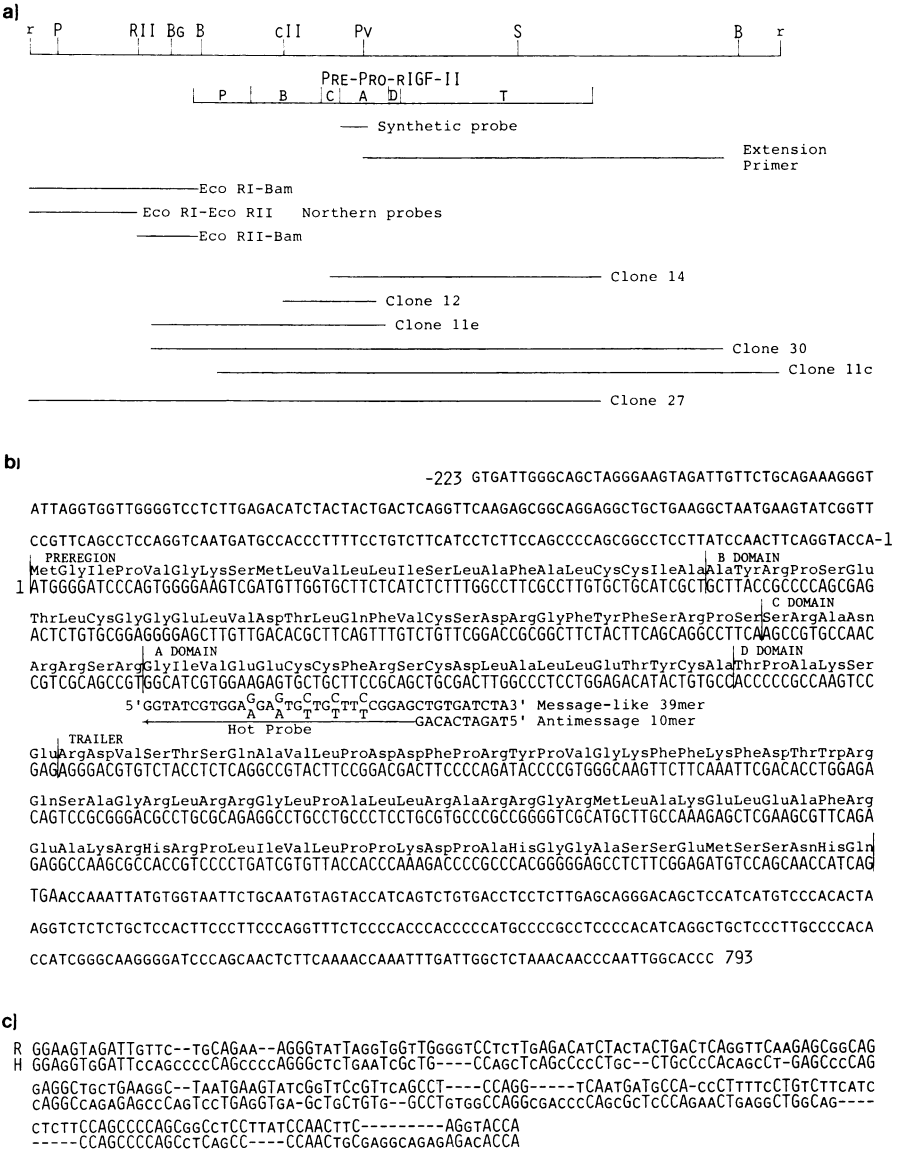

Fig. 2. a) Restriction map of the pre-pro-rIGF-II cDNA sequence shown in b). Only some of the unique restriction sites (P=Pst, Bg=Bgl I, B=Bam, cII=Hinc II, Pv=Pvu II, and S=Sac I) are indicated. The Eco RII site (RII), used for the generation of hybridization probes, is not unique. The region encoding the rIGF-II precursor (P=preregion, domains B, C, A and D of rIGF-II, and T= trailer polypeptide) is indicated in a second line. The extent of the sequence carried by the characterized, overlapping cDNA clones is shown. The ends of all inserts are flanked by Eco RI linkers. The 5' linker of clone 27 and the

3' linker of clone 11c are indicated by r in the restriction map. b) Composite of the DNA sequence derived from the six clones shown in a). Small and large capital letters represent the noncoding (5' and 3') and coding regions, respectively. The position and sequence of the synthetic DNA oligonucleotides used for the generation of a labeled screening probe are shown below the corresponding sequence at the beginning of the A domain. c) Alignment of the 5' noncoding regions of rat (R) and human (H) (22) pre-pro-IGF-II cDNAs. Homologous regions are indicated by large capital letters. Gaps were introduced to maximize homology.

rIGF-II (179 residues) consists of a signal peptide (23 residues), the previously recognized (12) rIGF-II sequence (67 residues; B domain 32, C domain 8, A domain 21, D domain 6), and a trailer polypeptide (89 residues), until the first terminator is encountered. This structure is the same as that of pre-pro-hIGF-II (18,22), and analogous to that of pre-pro-hIGF-I (21),despite the dissimilarities between the two human factors in the sequences of their preregions and trailer polypeptides. Since the significance of the trailer sequence (if any) is unknown, it cannot be argued for the moment that the pre-pro-IGFs are yet another example of polyproteins (31) with more than one functional parts.

The translated rIGF-II sequence is in agreement with the published protein sequence (12), except that the first residue of the C domain is a Ser (AGC), instead of a Gly (GGX). The assignment of a Gly in this position seems unequivocal from the protein sequencing data, while the fact that the nucleotide sequence is AGC (instead of GGC) has been firmly demonstrated by sequencing four independent clones. We feel that this excludes the possibility of reverse transcriptase errors. Thus, we do not know how to explain this difference, which is certainly not due to polymorphism, since the source of both the protein and the RNA was the same. We note that the residue in this position of the hIGF-II sequence (8,18,22) is also a Ser.

The calculated MW of the putative precursor is 19,968 daltons (preregion 2,330, rIGF-II 7,512, trailer 10,126). This size is consistent with a MW of 21 ± 1 K reported from the results of in vitro translation of fractionated BRL-3A RNA (32), which was designated as the "22 K" precursor. However, there is a significant discrepancy between the size of the 3.4 kb mRNA we detect and the reported size (12 S) of the translated fraction (32). Moreover, not all of the expected sizes of the processing products of pre-pro-rIGF-II, inferred from our sequencing data, correlate with the molecular species that have been observed in BRL-3A cells in vivo (33). From the translated DNA sequence, one would simply predict that the signal peptide is eliminated first, yielding a pro-protein of 17.6 K, which would then be processed to

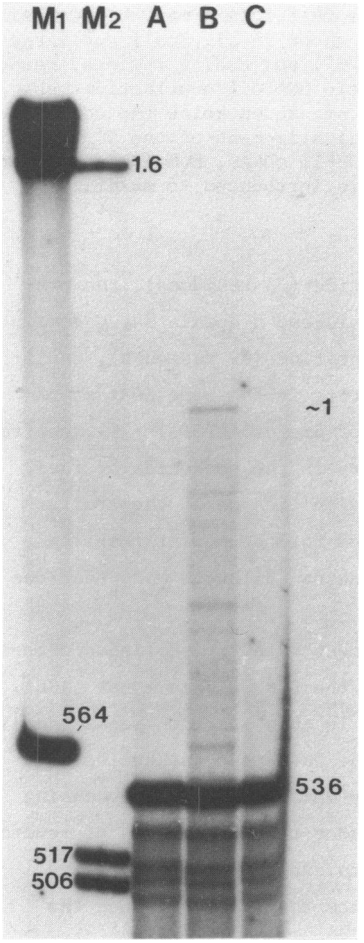

Fig. 3. Extension of a 536 nt antimessage primer (shown in lane A), synthesized from the M13 clone 30 (see Fig.2a), and extending from the end of the insert (5' end) to a Pvu II site at position +222 of the coding region (3' end). The RNA templates for reverse transcriptase were: 3.5 µg of poly $(A)^+$ RNA from BRL cells (lane B) and 10 µg of tRNA (control, lane C). Lanes M1 and M2 (labeled DNA markers) are Hind III fragments of phage λ and Hinf fragments of pBR322, respectively.

mature rIGF-II by clipping of the trailer polypeptide. However, the "22 K" in vitro precursor was not observed in vivo (33), even after a brief (10 min) ^{35}S-Met pulse, while a 19-20 K species (thought to represent the pro-protein) was the first detectable product. Also, an intermediate species of 16 K, and two smaller forms of 8.7 and 7.1 K,were detected in vivo. The latter two species were also present in the culture media. Since the 7.1 K species is presumably the 7.5 K mature rIGF-II, a possible interpretation of these observations is that the trailer polypeptide is not separated from the body of rIGF-II in one step. This possibility is not unlikely because the trailer sequence contains one Lys-Arg and two Arg-Arg pairs, which are candidates for processing sites. If this scheme turns out to be correct, it would be of

interest to examine the possible biological function of the small peptides that can be generated from the trailer. An alternative interpretation of these results (33) would be that the polyclonal antibody used in these studies recognizes also rIGF-II-related peptides.

The DNA sequencing data we present are only in partial agreement with the data by Dull et al. (18) and Whitfield et al. (19), who have also sequenced rIGF-II cDNA clones derived from the same mRNA source. Our data for the coding region of pre-pro-IGF-II are in complete agreement with the sequence of Dull et al. (18), including the first codon of the C domain (Ser instead of Gly). Thus, the accuracy of the sequence by Whitfield et al. (19), which has differences in six codons, should be re-examined. These authors presented the sequence of a partial cDNA clone, beginning with the last three codons of the C domain. They then extended this sequence upstream for another 35 codons by sequencing a primer-extended cDNA. Thus, their total sequence begins with the third residue of the B domain, and differs from ours in the third position of codons 3, 10, 22, 32 and 45, and in the first position of codon 33 (first codon of the C domain), which encodes Gly instead of Ser.

Our partial sequence of the 3' noncoding region covers 250 nt, and is in complete agreement with the longer sequence (392 nt) by Whitfield et al (19). The sequence by Dull et al. (18) is shorter in this region (134 nt). The last 16 nt of their sequence do not match our data.

The major discrepancy between our data and the sequence by Dull et al.(18) concerns the 5' noncoding region. With the exception of 9 nt upstream from the ATG initiator (where these authors position an intron in the corresponding sequence of the hIGF-II chromosomal gene), our sequence differs completely from theirs. The 5' noncoding region they present, derived from only one clone (pMSA-52H9), is 1055 nt long, which is inconsistent with the interpretation of our primer extension data. Moreover, an analogous primer extension experiment by Whitfield et al. (19) yielded a size estimate for this region, which is much closer to ours. We believe that our sequence for the 5' noncoding region of the pre-pro-rIGF-II mRNA is correct for the following reasons: a) Northern analysis (see below) indicates that this sequence is present in the 3.4 kb mRNA. b) The sequence reported by Bell et al. (22) as the 5' noncoding region of hIGF-II cDNA cannot be found in the hIGF-II gene sequence presented by Dull et al. (18), even though it can be aligned to our rIGF-II 5' noncoding region sequence (Fig.2c), despite the expected considerable degree of divergence.

Transcripts Related to Pre-Pro-rIGF-II

Since our sequence for the 5' noncoding region of the pre-pro-rIGF-II mRNA

Fig. 4. Alignment of the sequences of clones 19f, 27 (pre-pro-rIGF-II) and 19b. Large capital letters indicate homologous sequences. For details see text. The insert of clone 19b (approximately 230 nt) has been only partially sequenced by the enzymatic method because of the presence of a very strong stop immediately downstream of the displayed sequence.

was derived exclusively from clone 27 (Fig.2a), it was important to show that it did not correspond to a cloning artifact, but to a region of the 3.4 kb species. For this reason, we used as a probe a nick-translated DNA fragment from clone 27, extending from the extreme 5' end of the available sequence (Eco RI linker used for cloning) to an Eco RII site at position -79 (145 nt representing exclusively 5' noncoding region, Fig.2a). Northern analysis of BRL-3A RNA (Fig.1, lane C) showed that this sequence is indeed part of the 3.4 kb mRNA. Surprisingly, however, a second 1.6 kb RNA species was hybridized.

Thinking, after this result, that we might be dealing with a case of differential splicing, we rescreened the BRL-3A cDNA library with a 250 nt Eco RI-Bam probe (Fig.2a), which contains the entire available 5' noncoding region, and identified approximately 100 positive clones. Fifteen of these clones, selected at random, were first established as true-positives by rescreening with the same probe; they were then analyzed by Southern blotting (not shown), using as probe clone 14 DNA, which carries rIGF-II coding region (Fig.2a). Two of the clones (19f and 19b) which did not hybridize to the clone 14 probe, were sequenced (Fig.4). Clone 19f (a partial cDNA) has an insert of 400 nt, of which the last 51 nt are absolutely homologous to nucleotides -11 to -61 of our pre-pro-rIGF-II sequence, while the sequence upstream from position -61 is totally different. The insert of clone 19b is shorter (approximately 230 nt). It is also a partial sequence beginning only one nucleotide upstream from the breaking point of clones 19f and 27. Thus, we do not know whether 19f and 19b are overlapping clones. The sequence of clone 19b is completely homologous to the pre-pro-rIGF-II sequence between positions -61 and +53 (which includes the first 18 codons of the preregion). The rest of

the available sequence (73 nt) extending in the 3' direction diverges from the sequence of pre-pro-rIGF-II.

To examine whether the sequence of clone 19f was part of the 1.6 kb RNA, we made a probe from this clone that excluded the 19f/27 overlap, and analyzed by Northern blotting BRL-3A RNA (Fig.1, lane D). The result was unexpected: instead of the 1.6 kb RNA, another RNA species with a size of 1.75 kb was hybridized. However, when the 250 nt Eco RI-Bam probe from clone 27 was used as a probe, which includes the overlap of clones 19f, 19b and 27, not only the previously identified three RNA species (3.4, 1.75 and 1.6 kb) were hybridized, but also a fourth species of 1.1 kb gave a positive signal (Fig.5a, lane BRL). Hybridization to the Eco RII-Bam fragment of clone 27 shows that only three of the RNAs (3.4, 1.75 and 1.1 kb) share this sequence, because the 1.6 kb species did not hybridize (Fig.5b, lane A). The cross-hybridization properties of the various RNA species are summarized in Table 1. Three of these related RNAs (3.4, 1.75 and 1.1 kb) are polyadenylated species (see legend to Fig.1). Thus, in addition to the 3.4 kb pre-pro-rIGF-II mRNA, the 1.75 and 1.6 kb RNAs almost certainly also encode polypeptides. We do not know if the same applies for the 1.1 kb species, because its presence was examined only in total RNA.

In order to verify that the four related RNA species are not present exclusively in the transformed BRL-3A cell line, and in order to examine their possible developmental and tissue specificity, we hybridized total RNA from rat neonatal and adult tissues to the Eco RI-Bam probe of clone 27. Fig.5a shows that all of the neonatal (day 2 rat) tissues that were examined (brain, heart, liver, lung and muscle) contain all four transcripts, but in different amounts. (The neonatal muscle contains in addition a hybridizing band of 1.85 kb that appears unique to this tissue). In contrast, the adult liver and muscle tissues contain only RNA in the region of the 1.75/1.6 kb doublet (the two RNAs did not resolve well in this experiment). The developmental specificity of the 3.4 kb mRNA seems to be the same in the mouse, since a cross-hybridizing 3.4 kb transcript is present in fetal, but not in adult, mouse liver RNA (Fig.5c, lanes B and C). The absence of the 3.4 kb RNA from the adult tissues is consistent with previous data supporting the notion that IGF-II is a fetal somatomedin (see ref.6 and other references therein).

The differential expression of these transcripts in neonatal tissues is intriguing. Obviously muscle is the tissue with the highest concentration of the 3.4 and 1.1 kb species, while the 1.75 and 1.6 kb RNAs are underepresented. The latter two species, however, predominate in brain and lung. It is interesting that the cardiac muscle has a profile distinct from that of skeletal

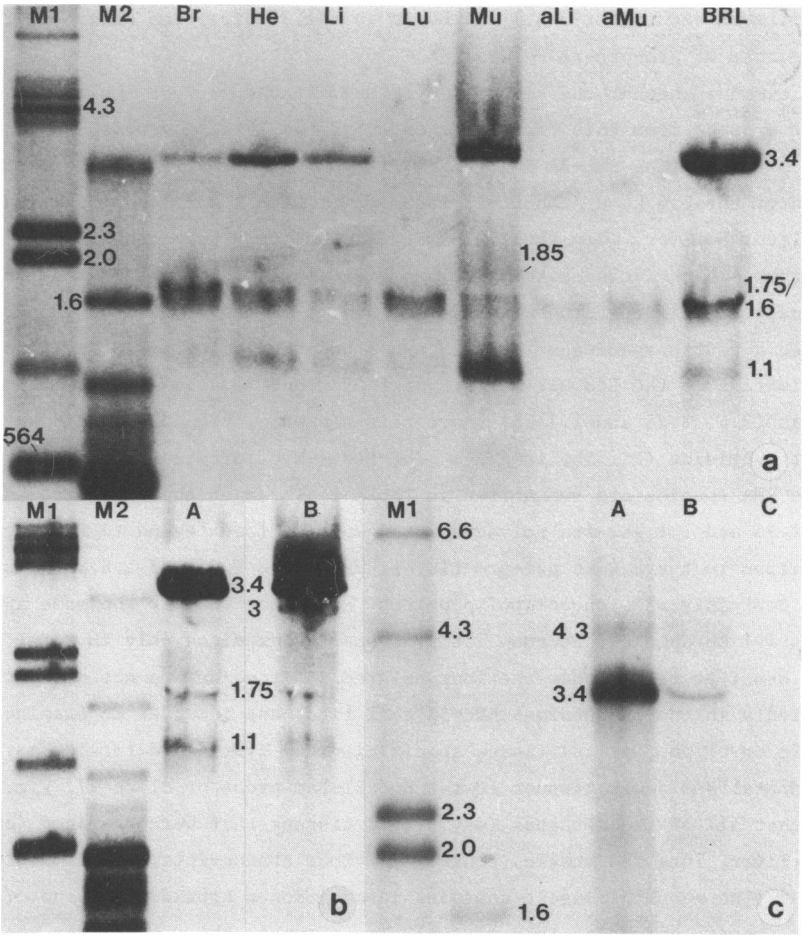

Fig. 5. a) Northern analysis of total RNA (50 µg per lane) extracted from neonatal rat tissues (lanes Br=brain, He=heart, Li=liver, Lu=lung, and Mu=muscle) or from adult tissues (lanes aLi=adult liver, and aMu=adult muscle) or from BRL-3A cells (lane BRL). The probe was uniformly labeled antimessage single-stranded DNA corresponding to the Eco RI-Bam fragment of clone 27 (see Fig. 2a). b) Two aliquots (50 µg each) of total BRL-3A cell RNA were electrophoresed in parallel and hybridized after transfer either to a uniformly labeled probe (lane A) corresponding to the Eco RII-Bam fragment of clone 27 (Fig.2a) or to a nick-translated probe (lane B) of the entire chromosomal clone hybridizing to rIGF-II sequences (see Materials and Methods and text). c) Hybridization of 8 µg of poly(A)$^+$ RNA from BRL cells (lane A), and 50 µg (each) of total RNA from mouse fetal liver (lane B) or adult liver (lane C). The probe was derived from clone 27 by limited synthesis (see legend to Fig.1) and represented primarily trailer polypeptide sequence. The size markers M1 and M2 are as in Fig. 3. Electrophoresis and transfer were as in Fig.1.

Table 1. Cross-hybridization Properties of Pre-pro-rIGF-II-related RNAs

Probe sequence	Pre-pro-rIGF-II 3' noncoding region	5' noncoding region			trailer region		
Clone	11c	27	27	27	27	19f	Genomic
Fragment		RI-B	RI-RII	RII-B		RI-MboII	
RNA (kb)							
4.3					+		
3.4(IGF)	+	+	+	+	+		+
1.75		+		+		+	+
1.6		+	+				
1.1		+		+			+

muscle. It is also interesting that the hepatic BRL-3A cell line retains the profile of the four species as it appears in liver tissue.

What is the source of these related RNA species? The available data do not allow us to discriminate between expression of different genes sharing exons (or parts of exons) and differential splicing. Preliminary Southern analysis of rat chromosomal DNA (not shown) yielded results compatible with the presence of a unique gene, when coding region of rIGF-II was used as a probe. However, additional bands (which for the moment cannot be accounted for) appeared after hybridization with the 250 nt Eco RI-Bam probe of clone 27. Nevertheless, the identity of the 27/19f or 27/19b overlapping sequence cannot be easily explained by the presence of more than one genes, because at least some minor degree of divergence should have been evident. On the other hand, the 27/19b homology cannot be explained by a simple model of differential splicing, because the two sequences do not share an exon, but parts of two exons, assuming that the rat and human (18) genes have the same intron organization. This issue (more than one gene vs differential splicing, or both) will be resolved from the characterization of a corresponding chromosomal gene that we have isolated. Though the exact representation of the 3.4 kb mRNA on this clone is still unknown, we note that a probe of the entire clone (13.2 kb insert) hybridizes to all of the RNAs, with the exception of the 1.6 kb species, and in addition to a 3.0 kb RNA (Fig. 5b, lane B). Though the latter species might belong to a co-ordinately expressed linked gene, the family of related RNAs does include additional members: a probe from clone 27 (representing primarily trailer sequence) hybridizes not only to the 3.4 kb mRNA, but also to a 4.3 kb transcript (Fig.5c, lane A).

These observations suggest that our data and the data by Dull et al. (18) might eventually be reconciled. In this regard, we note that differential spli-

cing is a likely explanation for the surprising (and uprecedented) observation by these authors that more than 1 kb of sequence, which they consider as 5' noncoding region, is 80% homologous between the human and rat IGF-II clones. We note that both of these sequences contain long open reading frames, which have the potential to encode proline-rich peptides (with some striking, but limited, homologies to retroviral gag proteins). In addition, differential splicing is consistent with the presence of an open reading frame in the available 3' noncoding region of the pre-pro-rIGF-II sequence. Alternatively, the 3.4 kb species might simply belong to a category of mRNAs with exceedingly long 3' noncoding regions, like, for example, the message encoding acetylcholine receptor (34).

Considering the questions generated from these data, we feel that it is premature to derive conclusions about the possible relationship between the IGF, insulin and relaxin genes, though the structure of the coding regions and the position of introns in the three gene types are not inconsistent with the postulated divergent (rather than convergent) mode of evolution (1,18). However, the most exciting prospect for the immediate future is the study of the developmental and tissue-specific expression of the structurally related transcripts we observe, especially if they are also functionally related, and the response of the corresponding gene(s) to hormonal stimuli. In addition, it will be interesting to eventually examine whether the genes encoding IGFs or their receptors, or both, correspond to oncogenes, by analogy to the PDGF/c-sis and EGF receptor/v-erb relationships (35-37).

ACKNOWLEDGEMENTS

We thank Matthew Rechler for providing the BRL-3A cell line, Tom Sargent for making available the rat chromosomal DNA library, Scott Zeitlin and Drew Murphy for advice and discussions, Sean Cotter for technical assistance, Vera Soares for help with the photography, and Eric Schon for critical reading of the manuscript. This work was supported by grants from the NIH to A.E. and D.I., and by a gift to A.E.'s laboratory from the Bristol-Myers company. M.S. was supported by a fellowship from the Brazilian Conselho Nacional do Desenvolvimento Cientifico e Tecnologico (CNPQ).

REFERENCES

1. Blundell, T.L. and Humbel, R.E. (1980) Nature 287, 781-787.
2. Smith, E.L. (1983) in Growth and Maturation Factors, Guroff, G. Ed., vol. 1, pp. 293-323, Wiley and Sons, New York.
3. Herington, A.C., Cornell, H.J., and Kuffer, A.D. (1983) Int.J.Biochem. 15, 1201-1210.
4. Preece, M.A. (1983) in Hormones in Blood, Gray, C.H. and James, V.H.T. Eds., vol 4, pp. 87-108, Academic Press, New York.
5. Scheonle, E., Zapf, J., Humbel, R.E., and Froesch, E.R. (1982) Nature 296, 252-253.

6. Adams, S.O., Nissley, S.P., Handwerger, S., and Rechler, M.M. (1983) Nature 302, 150-153.
7. Rinderknecht, E. and Humbel, R.E. (1978) FEBS Lett. 89, 283-286.
8. Rinderknecht, E. and Humbel, R.E. (1978) J.Biol.Chem. 253, 2769-2775.
9. Spencer, E.M., Ross, M., and Smith, B. (1983) in Insulin-like Growth Factors/Somatomedins, Spencer, E.M. Ed., pp. 81-96, de Gruyter, New York.
10. Rubin, J.S., Mariz, I., Jacobs, J.W., Daughaday, W.H., and Bradshaw, R.A. (1982) Endocrinology 110, 734-740.
11. Dulak, N.C. and Shing, Y.W. (1976) J.Cell Physiol. 90, 127-138.
12. Marquardt, H., Todaro, G.J., Henderson, L.E., and Oroszlan, S. (1981) J.Biol.Chem. 256, 6859-6865.
13. Dayhoff, M.O. (1978) Atlas of Protein Sequence and Structure, vol. 5 Supp. 3, p.151, Natl.Biomed.Res.Found., Washington, D.C.
14. Lomedico, P., Rosenthal, N., Efstratiadis, A., Gilbert, W., Kolodner, R., and Tizard, R. (1979) Cell 18, 545-558.
15. Perler, F., Efstratiadis, A., Lomedico, P., Gilbert, W., Kolodner, R., and Dodgson, J. (1980) Cell 20, 555-566.
16. Chan, S.J., Episkopou, V., Zeitlin, S., Karathanasis, S.K., MacKrell, A., Steiner, D.F., and Efstratiadis, A. (1984) Proc.Nat.Acad.Sci. USA 81, 5046-5050.
17. Episkopou, V., Murphy, A.J.M., and Efstratiadis, A. (1984) Proc.Nat.Acad. Sci. USA 81, 4657-4661.
18. Dull, T.J., Gray, A., Hayflick, J.S., and Ullrich, A. (1984) Nature 310, 777-781.
19. Whitfield, H.J., Bruni, C.B., Frunzio, R., Terrell, J.E., Nissley, S.P., and Rechler, M.M. (1984) Nature 312, 277-280.
20. Ullrich, A., Berman, C.H., Dull, T.J., Gray, A., and Lee, J.M. (1984) EMBO J. 3, 361-364.
21. 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.
22. Bell, G.I., Marrywhether, J.P., Pescador-Sanchez, R., Stempein, M.M., Priestley, L., Scott, J., and Rall, L.B. (1984) Nature 310, 775-777.
23. Maniatis, T., Fitsch, E.F., and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Sping Harbor, New York.
24. Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J., and Rutter, W.J. (1979) Biochemistry 18, 5294-5299.
25. Benton, W.D. and Davis, R.W. (1977) Science 196, 180-182.
26. Zeitlin, S. and Efstratiadis, A. (1984) Cell 39, 589-602.
27. Hu, N. and Messing, J. (1982) Gene 17, 271-277.
28. Sanger, F., Nicklen, S., and Coulson, A.R. (1977) Proc.Nat.Acad.Sci. USA 74, 5463-5467.
29. Maxam, A.M. and Gilbert, W. (1980) Meth.Enzymol. 65, 499-560.
30. Steiner, D.F., Quinn, P.S., Chan, S.J., Marsh, J., and Tager, H.S. (1980) Ann.N.Y.Acad.Sci. 343, 1-16.
31. Douglass, J., Civelli, O., and Herbert, E. (1984) Ann.Rev.Biochem. 53, 665-715.
32. Aquaviva, A.M., Bruni, C.B., Nissley, S.P., and Rechler, M.M. (1982) Diabetes 31, 656-658.
33. Yang, Y.W.-H., Acquaviva, A.M., Bruni, C.B., Romanus, J.A., Nissley, S.P., and Rechler, M.M. (1983) in Insulin-like Growth Factors/Somatomedins, Spencer, E.M. Ed., pp. 603-610, deGruyter, New York.
34. Noda, M., Furutani, Y., Takahashi, H., Toyosato, M., Tanabe, T., Shimizu, S., Kikyotani, S., Kayano, T., Hirose, T., Inayama, S., and Numa, S. (1983) Nature 305, 818-823.
35. Doolittle, R.F., Hunkapiller, M.W., Hood, L.E., Devare, S.G., Robbins,

K.C., Aaronson, S.A., and Antoniades, H.N. (1983) Science 221, 275-277.
36. Waterfield, M.D., Scrace, G.T., Whittle, N., Stroobant, P., Johnsson, A., Wasteson, A., Westermark, B., Hedlin, C.-H., Huang, J.S., and Deuel, T.F. (1983) Nature 304, 35-39.
37. Ullrich, A., Coussens, L., Hayflick, J.S., Dull, T.J., Gray, A., Tam, A.W., Lee, J., Yarden, Y., Libermann, T.A., Schlessinger, J., Downward, J., Mayes, E.L.V., Whittle, N., Waterfield, M.D., and Seeburg, P.H. (1984) Nature 309, 418-425.