Insulin-like growth factor I (IGF-I), a 70-amino acid basic polypeptide, plays a fundamental role in postnatal mammalian growth as a major mediator through which growth hormone exerts its biological effects. We have recently identified two human IGF-I cDNAs which predict distinct peptide precursors of 153 and 195 amino acids. In the present study, both cDNAs were used to isolate and characterize the human IGF-I gene from genomic libraries. The IGF-I gene extends over at least 45 kilobase pairs and contains five exons interrupted by four introns. The DNA sequence of exons 1 through 4 encodes the 195-amino acid precursor, while exons 1, 2, 3, and 5 code for the 153-residue peptide, confirming the hypothesis that at least two IGF-I mRNAs are generated by alternative RNA processing of the primary gene transcript. The structure of the IGF-I gene resembles that of its companion somatomedin, IGF-II, as judged by the analogous location of two introns and considerable nucleotide and amino acid sequence similarity, but appears more distantly related to other members of the insulin gene family. Restriction endonuclease polymorphisms in the IGF-I gene, which map near exon 5 as determined by Southern blot analysis, will be useful in defining the genetics of familial growth failure.

The somatomedins or insulin-like growth factors (IGFs'), a family of peptides which circulate in blood and stimulate DNA synthesis in a variety of cultured cells (1, 2), play important roles in mammalian growth and development. IGF-I, a 70-residue basic protein (3), is a major mediator of growth hormone action in postnatal life (1, 2, 4). Although the function of IGF-II, a 67-amino acid peptide (5), is less clear, it may serve an analogous role in fetal development (1).

While growth hormone may stimulate IGF-I biosynthesis in liver and possibly other tissues (6–9), only recent molecular cloning studies have begun to define the structure of IGF-I precursor peptides. From cDNA and gene cloning analyses (10–12), a single human IGF-I protein precursor was determined. We have documented a putative translation initiation site for the mRNA encoding this 153-amino acid peptide, IGF-IA, and have defined a cDNA coding for a second precursor of 195 residues, IGF-IB (13). Both protein precursors contain the 70-amino acid IGF-I found in serum but differ in the sequence and length of their carboxyl-terminal extensions. Since IGF-I is a single-copy gene (14, 15), these studies suggested that the two IGF-I cDNAs might represent alternatively spliced mRNA products arising from the gene transcript.

To investigate the biogenesis of the two IGF-I mRNA species, we have isolated the human IGF-I gene from bacteriophage genomic libraries using our cDNAs as probes and have determined the nucleotide sequence. Within its five exons the IGF-I gene encodes both mRNA species, confirming that at least two IGF-I mRNAs are produced by alternative RNA processing from a single primary transcript. As the distinct carboxyl extensions of each IGF-I precursor reside on separate exons, this study suggests that tissue-specific factors may play a role in IGF-I biosynthesis by influencing RNA splicing or protein processing and that the extension peptides themselves may subserve discrete biological functions.

EXPERIMENTAL PROCEDURES

Isolation of Recombinant Bacteriophage Containing the Human IGF-I Gene—A human genomic library was prepared using a modified λ 1059; λ Charon 30 (16, 17) hybrid vector, λ MG14, and size-fractionated human leukocyte DNA, partially digested with MboI (18). Approximately 4 x 10⁸ plaques from this library and another, derived from human fetal liver (19), were screened by standard methods (18) under Pl/EK1 containment using 3²P-labeled IGF-I cDNAs (13) as hybridization probes. DNA from plaque-purified positive isolates (20) was mapped using BamHI, EcoRI, and HindIII single and double digests by hybridization to 3²P-IGF-I cDNAs and human Ala probes (21). From 15 genomic clones containing IGF-I exons, 8 were selected for further analysis.

Nucleotide Sequence Analysis—Subclones containing each exon were prepared in plasmids pUC18 and 19 (22) for further restriction mapping and in M13 mp18 and 19 (22) for DNA sequencing using diisocy dye-chain-terminating inhibitors (23), 3²S-ATP, and both standard and gradient denaturing gel electrophoresis (24). Initial DNA sequence on exons 3 and 4 was obtained after preparing a series of overlapping deletions using Bal31 exonuclease (25), as indicated by the solid circles in the sequencing scheme. All other sequences were initiated at specific restriction endonuclease sites, as depicted by the short vertical lines in Fig. 1B. The arrows indicate the extent of the sequence determined. Except for 100 nucleotides of the intron preceding exon 2 and a portion of exon 4, all sequences presented in Fig. 2 were verified from both DNA strands, including all restriction endonuclease sites used as initiation points. The region of exon 4 sequenced in one orientation agrees completely with the corresponding cDNA (13).

Southern Blot Analysis—Ten micrograms of DNA from human

ALTHERNATIVE RNA PROCESSING PRODUCES TWO INSULIN-LIKE GROWTH FACTOR I PRECURSOR PEPTIDES

Organization and Sequence of the Human Insulin-like Growth Factor I Gene

(Received for publication, November 11, 1985)
AIGF-I Gene-Approximately 400,000 recombinant bacteriophage from each gene library were screened with the two IGF-I cDNA probes (IGF-IA encodes the 153-amino acid precursor and IGF-IB, the 195-residue peptide). From 15 positive clones, 8 were selected for detailed analysis by restriction endonuclease mapping and Southern blotting as depicted in Fig. 1. Five regions of DNA hybridized to the cDNA probes and have been designated exons 1–5. Exons 1, 2, 3, and 5 comprise IGF-IA mRNA and exons 1, 2, 3, and 4 IGF-IB mRNA (13). From the beginning of exon 1 to the end of exon 5, the gene extends for more than 45 kb. The exact size cannot be determined because a section of a large intron between exons 2 and 3 was not isolated from either gene library. Since this intron interrupts an asparagine codon in the B domain of IGF-I, it cannot contain an additional IGF-I exon. Within the IGF-I gene are 5 regions which hybridize strongly to middle repetitive DNA of the Alu type (21). Restriction enzymes include B, BamHI; E, EcoRl; and H, HindIII. As found in many other human genes, the Alu DNA maps to intervening and middle repetitive DNA of the Alu type (21). As the 5' end of exon 1 is not completely characterized, it has been left open. The vertical arrows under exons 4 and 5 mark the polyadenylation sites as determined by comparison with IGF-I cDNA sequences. The regions marked Alu signify hybridization to middle repetitive DNA of the Alu type (21). Restriction enzymes include B, BamHI; E, EcoRI; and H, HindIII. Below the restriction map are representative genomic clones. None of the λ isolates contains the portion of the intron between exons 2 and 3 depicted by the break in the restriction map, b, detailed restriction map of IGF-I exons 1 to 5 indicating DNA sequencing strategy. Boxes depict exons, coding regions are solid, and noncoding regions are cross-hatched.

RESULTS

Isolation of Recombinant Bacteriophage Containing the Human IGF-I Gene—Approximately 400,000 recombinant bacteriophage from each gene library were screened with the two IGF-I cDNA probes (IGF-IA encodes the 153-amino acid precursor and IGF-IB, the 195-residue peptide). From 15 positive clones, 8 were selected for detailed analysis by restriction endonuclease mapping and Southern blotting as depicted in Fig. 1. Five regions of DNA hybridized to the cDNA probes and have been designated exons 1–5. Exons 1, 2, 3, and 5 comprise IGF-IA mRNA and exons 1, 2, 3, and 4 IGF-IB mRNA (13). From the beginning of exon 1 to the end of exon 5, the gene extends for more than 45 kb. The exact size cannot be determined because a section of a large intron between exons 2 and 3 was not isolated from either gene library. Since this intron interrupts an asparagine codon in the B domain of IGF-I, it cannot contain an additional IGF-I exon. Within the IGF-I gene are 5 regions which hybridize strongly to middle repetitive DNA of the Alu type (21). As found in many other human genes, the Alu DNA maps to intervening and flanking sequences (21) as indicated on Fig. 1.

Nucleotide Sequence Analysis—The DNA sequencing strategy is depicted in Fig. 1 and the sequence is presented in Fig. 2. All exon-intron splice junctions and polyadenylation sites have been determined by comparison with IGF-I cDNA sequences (13). Although the transcription start site is not yet completely defined, by primer extension analysis using human liver RNA and a 32P-labeled oligonucleotide, the 5' end of IGF-I messenger RNA extends 583 nucleotides upstream of the priming site at position 1108 in Fig. 2 (data not shown). If the 5' untranslated region of the gene does not contain an intron, transcription must begin at position 525 or 526. Confirmation of the transcription start site awaits isolation of a cDNA clone containing additional 5' untranslated DNA and mapping of RNA by S1 nuclease. However, based on the primer extension study, cDNA, and gene sequence analyses, the 5' untranslated region is 684 nucleotides long. The largest cDNA which we have characterized extends 238 nucleotides 5' to the initiation codon (13) and contains several upstream in phase stop codons as denoted in Fig. 2. Of note also is a region of 42 nucleotides further 5' at positions 199–241 in the sequence containing alternating purine and pyrimidine residues with the potential to form Z DNA (29).

Exon 1 encodes 21 amino acids which may comprise the leader peptide, although no direct evidence exists to date on the site of initiation of translation of IGF-I. Exon 1 is followed by an intron of approximately 4.8 kb. Exon 2, 157 nucleotides, contains 52 amino acids and is identical to an IGF-I exon reported previously (11). The initial 27 residues precede the start of the B domain of IGF-I and contain 2 methionine residues which also might serve as translation start signals (30). Exon 2 also encodes the beginning 25 amino acids of the B domain of mature IGF-I. Codon 26, asparagine, is interrupted after its first base by a large intron of at least 21 kb. The remainder of the genetic information coding for mature IGF-I protein resides on exon 3; 11 nucleotides encoding the rest of the B domain, the 12 codons of the C domain, 21 of

---

2 P. Rotwein, unpublished observation.
Fig. 2. Sequence of the human IGF-I gene. DNA sequence was determined according to the scheme depicted in Fig. 1. Exons are in upper case, and introns and flanking DNA in lower case. Translation of the IGF-I precursor is indicated; the 70 amino acids of IGF-I are underlined. Figure 1...
the A domain, and 8 of the D region. In addition, exon 3 codes for 16 amino acids comprising a portion of the carboxyl-terminal extension or E domain and is followed by a splice junction.

Exons 4 and 5 each encode distinct E extension peptides, termination codons, 3' nontranslated regions, and poly(A) addition signals and sites. Exon 4, located 1.5 kb 3' to exon 5, consists of 515 nucleotides and comprises the 3' terminus of the putative IGF-IB mRNA. Exon 4 codes for a highly basic 61-amino acid extension peptide, with several dibasic and tribasic residues. The existence of this peptide is predicted solely from DNA sequence, as is the protein encoded by exon 5. Following the TGA termination codon are 329 bases of 3' nontranslated DNA. The polyadenylation signal, AATAAA, is found 16 nucleotides before the poly(A) addition site (31). Exon 5, located 17 kb downstream from exon 3 is 344 bases long. It contains 19 codons of the other basic extension peptide (see Fig. 2), a TAG terminator, and 284 bases of 3' nontranslated DNA. The polyadenylation signal AATAAA (31) precedes the poly(A) site by 19 nucleotides.

**DNA Polymorphisms Map to Exon 5 of the IGF-I Gene**—We have examined DNA from 18 unrelated adults of normal stature using IGF-IA and IGF-IB cDNA probes and 5 different restriction endonucleases. We find site polymorphisms for HindIII and PvuII (14), which map exclusively to exon 5 of the IGF-I gene and are linked to one another (Fig. 3). In a preliminary survey of DNA samples obtained from 8 growth-impaired individuals, a comparable frequency of polymorphic sites also is observed. Except for the HindIII and PvuII polymorphisms, the lengths of restriction fragments in genomic DNA which hybridize to IGF-I cDNA probes correspond precisely to those obtained from the recombinant bacteriophage clones, confirming that no DNA rearrangement occurred within the IGF-I gene during propagation of the gene libraries.

**IGF Genes and the Insulin Gene Family**—A comparison of the gene structures for IGF-I and IGF-II (32) supports their divergence from a common evolutionary ancestral gene (Fig. 4). Primary evidence includes DNA sequence identity of 61 and 49%, respectively, between IGF-I exons 2 and 3 and the corresponding regions of IGF-II, comparable amino acid sequence similarity (33), and the homologous location of two introns within each gene. Yet, compared with IGF-II, the human IGF-I gene has two additional exons, one potentially encoding the amino-terminal 21 residues of the protein precursor, and the other containing the alternative E domain. Even more dissimilar are the structures of the IGF-I and insulin or relaxin genes (34, 35), as evidenced by the disparate positions of introns within these genes.

**DISCUSSION**

In this paper we report the organization and nucleotide sequence of the human IGF-I gene. The gene contains several surprising features, including its unexpectedly large size and the presence of alternatively spliced exons. The IGF-I gene spans more than 45 kb on chromosome 12 (14, 15), its exons being separated by intervening sequences of 1.5 to greater than 21 kb. The exact length of the gene is unknown, as there is an uncloned gap of indeterminate size within the largest intron, which splits the B domain coding region, and as the 5' end is not yet completely mapped. Possibly the large size of the IGF-I gene impedes its transcription and mRNA processing, accounting for the low level of gene expression. This supposition is analogous to the correlation between the large size of the gene encoding another liver protein, blood clotting factor VIII, and its low level of biosynthesis (36). Further-
more, since IGF-I is required primarily to support linear growth, the levels of gene expression may vary during growth and development and are probably highest in actively growing animals. Alternative RNA splicing also may provide a point of regulation for IGF-I gene expression as has been suggested for other genes, such as calcitonin/calcitonin gene-related peptide (37), in which different combinations of exons function in different tissues. Since large IGF-I immunoreactive proteins have been detected in several tissues and cell lines (8, 38), these immunoreactive species may be the consequences of tissue-specific IGF-I mRNA processing and expression. In this context, further processing of the two different IGF-I protein precursors provides another potential level for control of IGF-I biosynthesis in which additional tissue factors may play a role. In addition, the two E peptides may serve specific, tissue-limited, biological functions. With the availability of IGF-I cDNA and genomic probes, the question of developmental regulation may now be studied, as may the related issues of regulation of IGF-I gene expression by growth hormone and other hormones and the determination of the sites of IGF-I mRNA biosynthesis. The functions of the alternatively expressed E regions may be examined using synthetic peptides and antibodies. The availability of probes also will spur investigation into the complexities of using synthetic peptides and antibodies. The availability of tissue factors may play a role. In addition, the two E peptides have been detected in several tissues and cell lines for preparing subclones of exon 5; John M. Chirgwin, Sharon Ogden, and Michael J. Maniatis, T., Fritsch, E. F., and Samboorok, J., (1982) in Molecular Cloning pp. 322–325, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Acknowledgments—We thank the following individuals: David Tie-meier for making possible the collaboration between P. R. and G. G. K.; Maynard Olsen for a gift of MG14 bacteriophage; Pranav Trivedi for preparing subclones of exon 5; John M. Chirgwin, Sharon Ogden, and David Tie-meier for critical comments on the manuscript; and Agnes Kelly for preparation of the manuscript.

REFERENCES